THE AMERICAN MATHEMATICAL

MONTHLY

Volume 106, Number 6 June–July 1999

AN OFFICIAL PUBLICATION OF THE MATHEMATICAL ASSOCIATION OF AMERICA

NOTICE TO AUTHORS

The MONTHLY publishes articles, as well as notes and other features, about mathematics and the profession. Its readers span a broad spectrum of mathematical interests, and include professional mathematicians as well as students of mathematics at all collegiate levels. Authors are invited to submit articles and notes that bring interesting mathematical ideas to a wide audience of MONTHLY readers.

The MONTHLY's readers expect a high standard of exposition; they expect articles to inform, stimulate, challenge, enlighten, and even entertain. MONTHLY articles are meant to be read, enjoyed, and discussed, rather than just archived. Articles may be expositions of old or new results, historical or biographical essays, speculations or definitive treatments, broad developments, or explorations of a single application. Novelty and generality are far less important than clarity of exposition and broad appeal. Appropriate figures, diagrams, and photographs are encouraged.

Notes are short, sharply focussed, and possibly informal. They are often gems that provide a new proof of an old theorem, a novel presentation of a familiar theme, or a lively discussion of a single issue.

Articles and Notes should be sent to the Editor:

ROGER A. HORN
1515 Mineral Square, Room 142
University of Utah
Salt Lake City, UT 84112

Please send your email address and 3 copies of the complete manuscript (including all figures with captions and lettering), typewritten on only one side of the paper. In addition, send one original copy of all figures without lettering, drawn carefully in black ink on separate sheets of paper. Authors who use LaTeX are urged to use article.sty and its standard environments with no custom formatting

Letters to the Editor on any topic are invited; please send to the MONTHLY's Utah office. Comments, criticisms, and suggestions for making the MONTHLY more lively, entertaining, and informative are welcome.

See the MONTHLY section of MAA Online for current information such as contents of issues and descriptive summaries of forthcoming articles:

http://www.maa.org/

Proposed problems or solutions should be sent to:

DANIEL ULLMAN, MONTHLY Problems
Department of Mathematics
The George Washington University
2201 G Street, NW, Room 428A
Washington, DC 20052

Please send 2 copies of all problems/solutions material, typewritten on only one side of the paper.

Reprint permission:
DONALD ALBERS, Director of Publication

Advertising Correspondence:
Dave Riska, *driska@maa.org*

Change of address, missing issues inquiries, and other subscription correspondence:
MAA Service Center, *maahq@maa.org*

All at the address:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036

Recent copies of the MONTHLY are available for purchase through the MAA Service Center, maahq@maa.org, 1-800-331-1622

Microfilm Editions: University Microfilms International, Serial Bid coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

Second class postage paid at Washington, DC, and additional mailing offices. ***Postmaster***: Send address changes to the American Mathematical Monthly, Membership / Subscription Department, MAA, 1529 Eighteenth Street, N.W., Washington, DC, 20036-1385.

Change of Variables in Multiple Integrals

Peter D. Lax

Dedicated to the memory of Professor Clyde Klipple, who taught me real variables by the R. L. Moore method at Texas A & M in 1944.

1. Let $y = \varphi(x)$ be a differentiable mapping of the interval $S = [c, d]$. Denote by T the interval $[a, b]$ with $\varphi(c) = a$, $\varphi(d) = b$. Let f be a continuous function of y. The change of variable formula says that

$$\int_S f(\varphi(x)) \frac{d\varphi}{dx}\, dx = \int_T f(y)\, dy. \tag{1.1}$$

The usual proof uses the fundamental theorem of calculus. Denote by g an anti-derivative of f:

$$f = \frac{d}{dy} g. \tag{1.2}$$

According to the fundamental theorem of calculus,

$$\int_T f(y)\, dy = g(b) - g(a), \tag{1.3}$$

where a and b are the endpoints of the interval T. On the other hand, by the chain rule the derivative of the composite $g \circ \varphi$ is given by

$$\frac{d}{dx} g(\varphi(x)) = \frac{dg}{dy} \frac{d\varphi}{dx}.$$

Using (1.2) we see that the x derivative of $g \circ \varphi$ is the integrand on the left in (1.1); therefore by the fundamental theorem of calculus,

$$\int_S f(\varphi(x)) \frac{d\varphi}{dx}\, dx = g(\varphi(d)) - g(\varphi(c)), \tag{1.4}$$

where c and d are the endpoints of the interval S. Since $\varphi(c) = a$ and $\varphi(d) = b$, the right sides of (1.3) and (1.4) are the same; this completes the proof of (1.1).

The usual proof of the change of variable formula in several dimensions uses the approximation of integrals by finite sums; see for instance [7]. The purpose of this note is to show how to use the fundamental theorem of calculus to prove the change of variable formula for functions of any number of variables. Then, as a surprising byproduct, we obtain a proof of the Brouwer fixed point theorem. In the last section we compare our proof with other known analytic proofs of the fixed point theorem.

I thank Daniel Velleman for suggesting a substantial simplification of the argument.

2. In this section we study mappings $\varphi(x) = y$ of n-dimensional x space into n-dimensional y space. We impose two assumptions:

i) φ is once differentiable.
ii) φ is the identity outside some sphere, say the unit sphere:

$$\varphi(x) = x \quad \text{for } |x| \geq 1.$$

Change of variable theorem:. *Let f be a continuous function of compact support. Then*

$$\int f(\varphi(x))J(x)\,dx = \int f(y)\,dy, \tag{2.1}$$

where J is the Jacobian determinant of the mapping φ:

$$J(x) = \det\frac{\partial\varphi_j}{\partial x_i}; \tag{2.2}$$

here φ_j is the j^{th} component of φ.

We prove this for functions f that are once differentiable and for mappings φ that are twice differentiable; since functions and mappings can be approximated by differentiable ones, this suffices. The approximation can be accomplished by mollification, that is, by convolving each component of φ with a smooth, nonnegative, spherically symmetric function m with small support whose integral equals 1. As the support of m shrinks to zero, $m * \varphi$ and its first derivatives tend to those of φ. If φ is the identity, so is $m * \varphi$.

Define

$$g(y_1, y_2, \ldots, y_n) = \int_{-\infty}^{y_1} f(z, y_2, \ldots, y_n)\,dz. \tag{2.3}$$

Clearly, $\frac{\partial g}{\partial y_1} = f$. Since f is once differentiable, so is g. Since f has compact support, we can choose c so large that f is zero outside the c-cube

$$|y_i| \leq c, \quad i = 1, 2, \ldots, n.$$

It follows from (2.3) that $g(y_1, \ldots, y_n) = 0$ when $|y_j| \geq c$ for any $j \neq 1$, and when $y_1 \leq -c$.

Take $c \geq 1$; then the c-cube contains the unit ball. Since φ is the identity outside the unit ball, $f(\varphi(x))$ is zero outside the c-cube in x-space. So in the integrals in (2.1) we may restrict integration to the c-cube.

In the left side of (2.1), express f as the partial derivative of g:

$$\int \frac{\partial g}{\partial y_1}(\varphi(x))J(x)\,dx. \tag{2.4}$$

We denote by D the gradient with respect to x; the columns of the Jacobian matrix $\partial\varphi/\partial x$ are $D\varphi_1, \ldots, D\varphi_n$.

Observation:. *The integrand in* (2.4) *can be written as the following determinant*:

$$\det(Dg(\varphi), D\varphi_2, \ldots, D\varphi_n). \tag{2.5}$$

Proof: By the chain rule

$$Dg(\varphi) = \sum_{j=1}^{n} (\partial_{y_j} g) D\varphi_j. \tag{2.6}$$

We set this into the first column in (2.5). Formula (2.6) expresses $Dg(\varphi)$ as a linear combination of the vectors $D\varphi_1, D\varphi_2, \ldots, D\varphi_n$; the last $n - 1$ of these vectors are the last $n - 1$ columns of the matrix in (2.5), and therefore these can be subtracted from $Dg(\varphi)$ without altering the value of the determinant (2.5). This leaves us with

$\det((\partial_{y_1} g(\varphi))D\varphi_1, D\varphi_2, \ldots, D\varphi_n)$; factoring out the scalar $(\partial_{y_1})g(\varphi)$ gives $(\partial_{y_1})g(\varphi)J$, the integrand in (2.4). ■

The next step is to expand the determinant (2.5) according to the first column; we obtain

$$M_1 \partial_{x_1} g(\varphi) + \cdots + M_n \partial_{x_n} g(\varphi), \tag{2.7}$$

where $M_1, \ldots, M_n$ are the cofactors of the first column of the Jacobian matrix. Setting (2.7) into the integrand in (2.4) we get

$$\int \big(M_1 \partial_{x_1} g(\varphi) + \cdots + M_n \partial_{x_n} g(\varphi)\big)\, dx. \tag{2.8}$$

Since φ is twice differentiable, we can integrate each term by parts over the c-cube and obtain

$$-\int g(\varphi)\big(\partial_{x_1} M_1 + \cdots + \partial_{x_n} M_n\big)\, dx + \textit{boundary terms}. \tag{2.9}$$

We use now the following classical identity:

$$\partial_{x_1} M_1 + \cdots + \partial_{x_n} M_n \equiv 0. \tag{2.10}$$

We sketch a proof: We can write the left side of (2.10) symbolically as

$$\det(D, D\varphi_2, \ldots, D\varphi_n). \tag{2.11}$$

For $n = 2$ we have

$$\det(D, D\varphi_2) = \partial_1 \partial_2 \varphi_2 - \partial_2 \partial_1 \varphi_2 = 0.$$

For $n > 2$ we note that the cofactors M_j are multilinear functions of the φ_j. Using the product rule of differentiation, we write (2.11), again symbolically, as

$$\sum_{2 \le k \le n} \det(D, D\varphi_2, \ldots, D\varphi_n)_k, \tag{2.12}$$

where the subscript k means that the differential operator D in the first column acts only on the k^{th} column. We leave it to the reader to verify that each of the determinants in the sum (2.12) is zero.

The identity (2.10) shows that the n-fold integral in (2.9) is zero.

We turn now to the boundary term in (2.9). Since $g(\varphi(x)) = g(x)$ on the boundary of the c-cube, the only nonzero boundary term is from the side $x_1 = c$; since $M_1 = 1$ when $\varphi(x) \equiv x$, that boundary term is

$$\int g(c, x_2, \ldots, x_n)\, dx_2 \cdots dx_n. \tag{2.13}$$

Using the definition (2.3) of g in (2.13) gives

$$\int\int_0^c f(z, x_2, \ldots, x_n)\, dz\, dx_2 \cdots dx_n,$$

which is the right side of equation (2.1). This completes the proof of the change of variables formula.

3. In our proof of the change of variables formula, we assumed neither that φ is one-to-one, nor that it is onto. We claim:

A mapping φ having properties i) and ii) of the change of variables theorem maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.

Suppose some point y_0 were not the image of any x. Since φ is the identity outside the unit ball, y_0 would lie inside the unit ball. Since φ maps $|x| \leq 1$ into a closed set, it would follow that some ball B_0 centered at y_0 would be free of images of φ. Now take any function f supported in the ball B_0, whose integral is nonzero:

$$\int f \, dy \neq 0. \tag{3.1}$$

By the change of variable formula

$$\int f(\varphi(x)) J \, dx = \int f \, dy \neq 0. \tag{3.2}$$

Since the range of φ avoids B_0, and since the support of f lies in B_0, the integrand on the left in (3.2) is identically zero; then so is the integral. This contradicts (3.1), and so the claim is established.

Intermediate Value Theorem:. *Let φ be a continuous map of the unit ball in $\mathbb{R}^n$ into $\mathbb{R}^n$ that is the identity on the boundary:*

$$\varphi(x) = x \quad \text{for } |x| = 1.$$

Then the image of φ covers every point in the unit ball.

Proof: Extend φ to be the identity outside the unit ball. Then approximate the extended map by differentiable maps, each the identity outside the unit ball. According to our claim, each of these maps covers the unit ball. By compactness, so does their limit. ■

The following well-known argument shows how to deduce the Brouwer fixed point theorem from the intermediate value theorem.

Let ψ be a continuous mapping of the unit ball into the unit ball; we claim that it leaves a point fixed. If not then for every x there is a ray from $\psi(x)$ through x. This ray pierces the unit ball at a point that we denote by $\varphi(x)$. Clearly, φ is a continuous mapping; it is the identity for x on the unit sphere and maps the unit ball into the unit sphere. This contradicts the intermediate value theorem. ■

4. The Brouwer fixed point theorem has many analytical proofs. How do they compare with the present one? Hadamard [**3**] employed the identity (2.10) about the Jacobian matrix; so did Dunford-Schwartz [**2**, pp. 467–470].

Samelson [**6**] used Stokes' theorem to give an extremely short proof of the Brouwer fixed point theorem. This proof was rediscovered by Kannai [**5**]. According to Laurent Schwartz, as related by Haim Brézis, such a proof was current in Paris in the thirties.

Báez-Duarte [**1**] proved formula (2.1) using exterior forms and Stokes' theorem and deduced from it the intermediate value theorem. My deduction is the same as Báez-Duarte's.

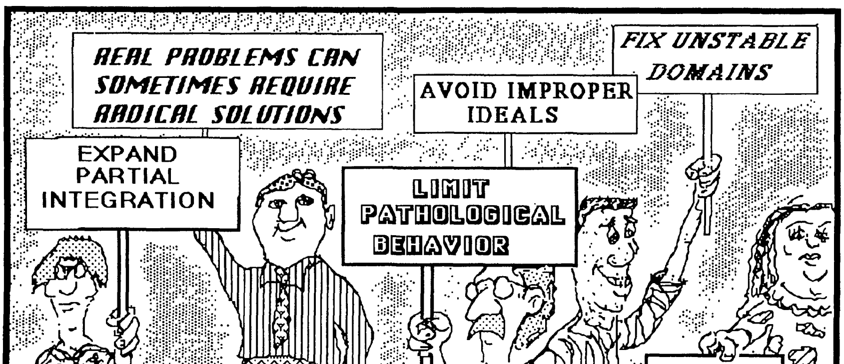
The integration of exterior forms over chains presupposes the change of variable formula for multiple integrals. It is amusing that the change of variables formula alone implies Brouwer's theorem.

In conclusion we call attention to Erhardt Heinz's beautiful analytic treatment of the Brouwer degree of a mapping.

REFERENCES

1. L. Báez-Duarte, Brouwer's Fixed-Point Theorem and a Generalization of the Formula for Change of Variables in Multiple Integrals, *J. Math. Anal. Appl.* **177** (1993) 412–414.
2. N. Dunford and J. Schwartz, *Linear Operators, Part I*, Wiley-Interscience, New York, 1958.
3. J. Hadamard, Sur quelques applications de l'indice de Kronecker, pp. 437–477 in J. Tannery, *Introduction à la théorie des fonctions d'une variable*, vol. 2, Paris, 1910.
4. E. Heinz, An Elementary Analytic Theory of the Degree of Mapping n-Dimensional Space, *J. Math. Mech.* **8** (1959) 231–248.
5. Y. Kannai, An elementary proof of the no retraction theorem, *Amer. Math. Monthly* **88** (1981) 264–268.
6. H. Samelson, On the Brouwer fixed point theorem, *Portugal. Math.* **22** (1963) 189–191.
7. J. Schwartz, The formula for change in variables in a multiple integral, *Amer. Math. Monthly* **61** (1954) 81–85.

PETER LAX was born in Hungary in 1926; he came to the U.S. in December, 1941 on the last boat. He is a fixture at the Courant Institute of New York University; his mathematical interests are too numerous to mention. He has always liked to teach at all levels, hence this paper.
Courant Institute, NYU, 251 Mercer St., New York, NY 10012
lax@cims.nyu.edu



Contributed by Russ Hood, Rio Linda, CA

Random Walks and Plane Arrangements in Three Dimensions

Louis J. Billera, Kenneth S. Brown, and Persi Diaconis

1. INTRODUCTION. The geometry of hyperplane arrangements in Euclidean space is a rich subject, which touches geometry [**14**], combinatorics [**21**], and operations research [**23**]. Probability was introduced into the subject by Bidigare, Hanlon, and Rockmore [**2**], who found a natural family of random walks associated with hyperplane arrangements. These walks were studied further by Brown and Diaconis [**6**]. One reason this development is exciting is that the walks admit a rather complete theory. We introduce the reader to this circle of ideas by specializing to the 3-dimensional case (planes in $\mathbb{R}^3$). Here we are able to use tools from geometry to obtain a surprising formula for the stationary distribution of the walk.

1.1. The geometric setup. Consider a collection $\mathscr{A}$ of n planes through the origin in $\mathbb{R}^3$. We assume throughout this paper that the intersection of the planes is $\{0\}$. In particular, $n \geq 3$. It is useful to picture the arrangement $\mathscr{A}$ via the intersection of the planes with the unit sphere S^2. See Figure 1 for the case where $\mathscr{A}$ consists

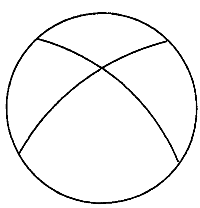


Figure 1. Three great circles on S^2.

of the three coordinate planes. The picture shows the Northern Hemisphere, viewed from above the North Pole; thus the outer circle is the equator $z = 0$. Adding the plane $z = x + y$ gives the picture in Figure 2.

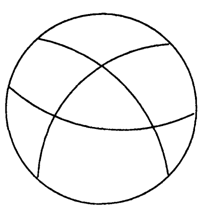


Figure 2. Four great circles on S^2.

The n great circles corresponding to the arrangement $\mathscr{A}$ decompose the sphere into cells: There are f_0 vertices, f_1 edges, and f_2 regions. When $n = 3$, for example, one has $f_0 = 6$, $f_1 = 12$, and $f_2 = 8$. Only the cells in the Northern Hemisphere are visible in the picture, but the Southern Hemisphere has exactly the same geometry. Notice that all the regions are triangular in Figure 1. In Figure 2 there are $f_2 = 14$ 2-dimensional cells: 8 triangles and 6 quadrilaterals, half of which are visible.

The cell decomposition induced by an arrangement has the following special property: *Given a region C and a vertex v, there is a unique region C' adjacent to v that is closest to C, in the sense that C' is separated from C by the minimum number of great circles*. The region C' is said to be the *projection* of C on v and is denoted vC. We explain in Section 3.2 why it is uniquely defined. Figure 3 shows an example; here C' is at distance two from C.

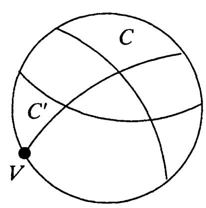


Figure 3. The projection of C on v.

1.2. A random walk on the regions. Bidigare, Hanlon, and Rockmore [**2**] used the projection operators to define the following walk on the regions: If the walk is in region C, choose a vertex v at random and move to the projection $C' = vC$. (We assume here that the vertices are chosen uniformly, so that all f_0 vertices are equally likely.) This walk is described mathematically by its *transition matrix* K. The rows and columns of K are indexed by the regions, with $K(C, C')$ being the chance of moving from C to C' in one step. Thus

$$K(C, C') = \frac{1}{f_0}\Lambda(C, C'),$$

where $\Lambda(C, C')$ is the number of vertices v of C' such that $vC = C'$.

Suppose, for instance, that $\mathscr{A}$ consists of the three coordinate planes. The 8 triangular regions are the intersections with the sphere of the 8 orthants in $\mathbb{R}^3$, as indicated in Figure 4. For example, the region $+ - +$ corresponds to the orthant $x > 0$, $y < 0$, $z > 0$. The matrix Λ is shown in Table 1; one has $K = \frac{1}{6}\Lambda$.

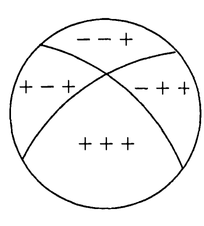


Figure 4. The regions correspond to orthants.

TABLE 1 The matrix Λ.

	$+++$	$++-$	$+-+$	$+--$	$-++$	$-+-$	$--+$	$---$
$+++$	3	1	1	0	1	0	0	0
$++-$	1	3	0	1	0	1	0	0
$+-+$	1	0	3	1	0	0	1	0
$+--$	0	1	1	3	0	0	0	1
$-++$	1	0	0	0	3	1	1	0
$-+-$	0	1	0	0	1	3	0	1
$--+$	0	0	1	0	1	0	3	1
$---$	0	0	0	1	0	1	1	3

Notice that for any region C, the three vertices of C satisfy $vC = C$; this explains the diagonal entries of Λ. Projection onto each of the remaining three vertices flips C to an adjacent region, thus accounting for the three 1's in each row. At each step, then, the walk stays in its current region with probability $1/2$ and otherwise moves to a randomly chosen adjacent region. This example is unusual, in that one never moves a distance of more than 1 in any step; typically the walk is much more vigorous.

The l^{th} power of the matrix K gives the transition probabilities after l steps; for example,

$$K^2(C,C') = \sum_{C''} K(C,C'')K(C'',C'),$$

which is the chance of moving from C to C' in two steps. After all, to get from C to C' in two steps the walk must go to some C'' and then to C'.

A fundamental theorem of Markov chain theory [**17**, Theorem 4.1.4] implies that $K^l(C,C')$ tends to a limit $\pi(C')$, independent of the starting region C:

$$K^l(C,C') \to \pi(C') \quad \text{as } l \to \infty.$$

(The theorem requires a mild regularity condition, which is satisfied by our chain.) Here π is a probability distribution on the set of regions, and $\pi(C')$ represents the chance that the random walk is in C' after a large number of steps from any starting region C. The distribution π is called the *stationary distribution* of the walk. It can be characterized as the unique probability distribution satisfying

$$\sum_{C} \pi(C)K(C,C') = \pi(C') \tag{1}$$

for all C'. This says that π, viewed as a row vector, is a left eigenvector for K with eigenvalue 1.

1.3. Analysis of the walk. Bidigare, Hanlon, and Rockmore [**2**] determined the eigenvalues of K, which turn out to be real (and even rational):

(a) 1 is an eigenvalue of multiplicity 1.
(b) For each plane $H \in \mathscr{A}$, there is an eigenvalue

$$\lambda_H = \frac{\#\text{ of vertices on } H \cap S^2}{f_0}$$

of multiplicity 1.
(c) $2/f_0$ is an eigenvalue of multiplicity $(f_2 - 2)/2$.
(d) 0 is an eigenvalue of multiplicity $(f_2 - 2)/2 - n + 1$.

A different proof of this can be found in [**6**], where it is shown further that K is diagonalizable. Yet another proof is given in [**5**]. For a simple example, take $n = 3$ again; the 8 eigenvalues are then $1, \frac{2}{3}, \frac{2}{3}, \frac{2}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0$.

A second result of [**2**, **6**] is a surprisingly simple estimate relating the eigenvalues to the rate of convergence to stationarity. Let K_C^l be the distribution of the walk started from C after l steps, i.e., $K_C^l(C') = K^l(C, C')$. We measure convergence rate by means of the distance $\|K_C^l - \pi\|$ defined as follows: If P and Q are probability distributions on a finite set X, then

$$\|P - Q\| = \max_{A \subseteq X} |P(A) - Q(A)|. \tag{2}$$

The version of the convergence rate estimate given in [**6**] is

$$\left\| K_C^l - \pi \right\| \leq \sum_{H \in \mathscr{A}} \lambda_H^l, \tag{3}$$

for any starting region C. When $n = 3$, for example, the distance to stationarity is bounded by $3\left(\frac{2}{3}\right)^l$ after l steps.

Finally, a method is given in [**6**] for computing the stationary distribution π, though it is not always easy to get a useful formula for π by this method. Our main result is an explicit formula for π in the 3-dimensional case:

Theorem 1. *Let $\mathscr{A}$ be an arrangement of n planes in $\mathbb{R}^3$ whose intersection is $\{0\}$. Let π be the stationary distribution of the random walk on regions of the sphere, as described above. If C is a region with i sides, then*

$$\pi(C) = \frac{i - 2}{2(f_0 - 2)}.$$

We find this result quite surprising. It is not even obvious to us why all i-gons should have the same stationary probability, let alone why this probability should be proportional to $i - 2$. In our $n = 3$ example, the theorem says that the stationary distribution is uniform: $\pi(C) = \frac{1}{8}$ for each of the 8 triangles C.

1.4. Organization of the paper. In Section 2 we make several remarks about the main theorem and illustrate it with examples. In Section 3 we explain how to describe vertices, edges, and regions by means of sign sequences; this description is used in the proof of our main theorem. Section 4 gives some background on hyperplane arrangements in order to put the results stated in Section 1.3 into a broader context. We also state in Section 4 some of the results of [**2**, **6**] that were specialized to the 3-dimensional case in Section 1.3. In Section 5 we show how the general theory specializes to card shuffling and random tiling models. We work out some 3-dimensional cases of this in detail. Sections 4 and 5 are included mainly for motivation; they are not needed for the proof of Theorem 1. The proof of the latter begins in Section 6, where we give a geometric description of the matrix K. With this description available, it is quite easy to complete the proof; we do this in Section 7. Finally, Section 8 contains some pointers to the literature on plane arrangements in $\mathbb{R}^3$.

The key arguments of this paper represent the combinatorial essence of a hyperplane arrangement as a collection of sequences of signs; see Section 3. This can be abstracted to the notion of an oriented matroid. The present project makes a nice introduction to these ideas.

2. EXAMPLES AND REMARKS

2.1. General position. The n planes are said to be in *general position* if no three of the planes have a nonzero intersection or, equivalently, if only two great circles pass through each vertex on the sphere. Many of the examples in this paper are in fact in general position; but the simplicial arrangements to be discussed in Section 2.3 are never in general position unless $n = 3$, nor are the arrangements in Section 5 below.

It is easy to count cells in the general position case. One finds

$$f_0 = n(n-1), \quad f_1 = 2n(n-1), \quad \text{and} \quad f_2 = n(n-1) + 2.$$

In fact, there are $\binom{n}{2}$ pairs of great circles, each determining a pair of antipodal vertices, whence the first equation. For the second equation, note that each of the n great circles is cut into $2(n-1)$ arcs by the other $n-1$ great circles. [Alternatively, count the vertex-edge pairs in two different ways to get $2f_1 = 4f_0$, so that $f_1 = 2f_0$.] Finally, the third equation can be proved by a straightforward inductive argument, or it can be deduced from the first two via Euler's relation $f_0 - f_1 + f_2 = 2$.

The list of eigenvalues λ and multiplicities m_λ given in Section 1.3 becomes

λ	m_λ
1	1
$2/n$	n
$1/\binom{n}{2}$	$\binom{n}{2}$
0	$\binom{n}{2} - n + 1$

And the convergence rate estimate (3) is

$$\left\| K_C^l - \pi \right\| \le n\left(\frac{2}{n}\right)^l.$$

This shows that the distance to the stationary distribution is small after $l = 2$ steps if n is large. In Section 6.3, after we have a geometric description of the transition matrix K, we describe a family $\mathscr{A}_n$ of arrangements of n planes in general position such that $\|K_C - \pi\| \ge c > 0$ for all n. In this sense the walk is close to stationary after 2 steps, but not after 1 in general.

Another feature of the general position case is that if one wants to carry out the random walk algorithmically, it is quite easy to pick a random vertex: First pick a pair of great circles at random, so that all $\binom{n}{2}$ are equally likely; this defines a pair of antipodal points. Now choose one of these two points with probability $\frac{1}{2}$.

Finally, Theorem 1 gives the following formula for the stationary distribution in the general position case:

$$\pi(C) = \frac{i-2}{2(n(n-1)-2)}$$

if C is an i-gon.

Here are some specific examples:

Example 1. For $n = 3$, the arrangement is combinatorially equivalent to that of Figure 1. There are 8 triangular regions, and the stationary distribution is $\pi(C) = \frac{1}{8}$ for all C.

Example 2. For $n = 4$ planes in general position (Figure 2) there are 14 regions: 8 triangles with stationary probability $\pi(C) = \frac{1}{20}$ and 6 quadrilaterals with $\pi(C) = \frac{2}{20}$.

Example 3. With $n = 5$ planes in general position, the situation is as in Figure 5 up to isomorphism. There are 10 triangles with $\pi(C) = \frac{1}{36}$, 10 quadrilaterals with $\pi(C) = \frac{2}{36}$, and 2 pentagons with $\pi(C) = \frac{3}{36}$.

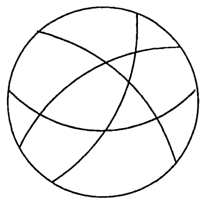


Figure 5. Five great circles in general position.

Example 4. With $n = 6$ or more planes the number of regions of each type can vary with the actual arrangement. From the catalogue in Grünbaum **[12, 14]** one sees that up to combinatorial equivalence there are exactly four general position arrangements of 6 planes, as shown in Figure 6. The number of i-gons ($i = 3, 4, 5, 6$)

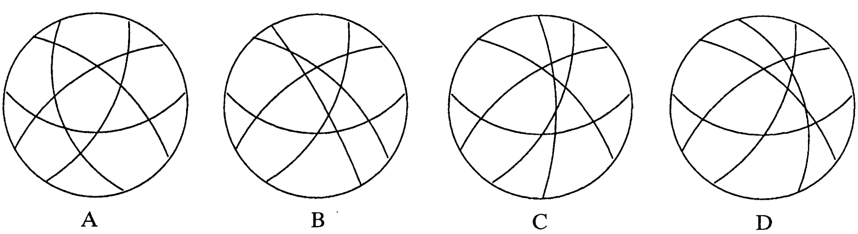


Figure 6. Six great circles in general position.

for each of these four arrangements is shown in Table 2. Each arrangement has $n(n-1) + 2 = 32$ regions. For each arrangement, the stationary probability of an i-gon is $\frac{i-2}{56}$. Note that, for some values of i, some arrangements don't have any i-gons. Nonetheless, the stationary probabilities sum to 1.

TABLE 2 The number of i-gons.

i	A	B	C	D
3	20	14	12	12
4	0	12	16	18
5	12	6	4	0
6	0	0	0	2

2.2. The Euler relation. For a given arrangement of n planes, let p_i be the number of i-sided regions. Since the stationary probabilities in Theorem 1 must sum to 1, we get

$$\sum_{i=3}^{n} p_i \frac{(i-2)}{2(f_0-2)} = 1. \tag{4}$$

We now show that this relation is equivalent to the Euler relation $f_0 - f_1 + f_2 = 2$. Rewrite (4) as $\Sigma i p_i - 2\Sigma p_i = 2(f_0 - 2)$. The first sum is the number of pairs consisting of a region and an edge of that region; since each edge is on exactly two regions, this sum equals $2f_1$. The second sum is f_2, so (4) becomes $2f_1 - 2f_2 = 2(f_0 - 2)$, which is Euler's identity.

2.3. Simplicial arrangements. The arrangement $\mathscr{A}$ is said to be *simplicial* if every region on the sphere is a triangle. The simplicial case has long held a special fascination for geometers [**13**, **14**]. And it is of special interest for the probability story also. In fact, every arrangement has some triangles (at least n of them by Levi's theorem [**14**, p. 25]); so Theorem 1 has the following consequence:

Proposition 1. *The random walk on regions has a uniform stationary distribution if and only if the arrangement is simplicial.*

There are three known infinite families of simplicial arrangements and 90 "sporadic" examples. It has been conjectured that there are no more infinite families and at most finitely many additional sporadic examples; see [**14**, p. 8, Conjecture 2.1]. A catalogue showing the first 89 sporadic examples can be found in [**13**], and the 90th is in [**14**]. (Note: It is stated in [**14**] that there are 91 sporadic examples, but the arrangements called $A_2(17)$ and $A_7(17)$ in [**13**] have been shown to be isomorphic; see [**15**, p. 59].) Many of the examples have a great deal of symmetry, and one is therefore not surprised when the stationary distribution turns out to be uniform. But there is one known example ([**14**, pp. 8–9]) of a simplicial arrangement whose group of combinatorial symmetries is trivial. Nevertheless, the walk on the regions has a uniform stationary distribution. This example, which has $n = 28$, is shown in Figure 7 in its projective representation. We discuss such

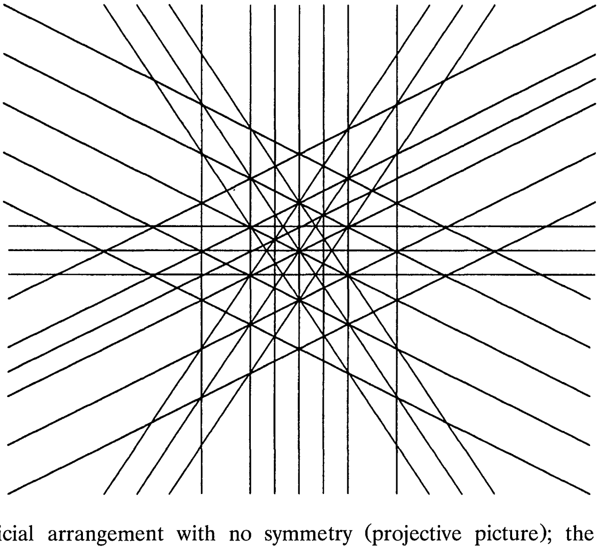


Figure 7. A simplicial arrangement with no symmetry (projective picture); the line at infinity is included.

representations in Section 8; for now, suffice it to say that we get the spherical representation by drawing the lines in Figure 7 as great semicircles on the Northern Hemisphere. Parallel lines yield great semicircles with the same endpoints on the equator.

2.4. Arrangements of pseudocircles. The theory developed in [**6**] and in the present paper goes through if the family of great circles is replaced by a family of "pseudocircles," i.e., simple closed curves on the sphere that are not necessarily great circles but merely intersect as great circles would. See [**4**, **14**, **23**] for the precise definition and many examples. The reason the theory goes through is that, as we remarked in the introduction, all of our work is based on the oriented matroid underlying the arrangement, and this exists for arrangements of pseudocircles also. Generalizing in this way is of interest because it vastly increases the supply of examples. In particular, one obtains seven new infinite families of simplicial arrangements; see [**14**, **11**].

3. AN ENCODING OF THE CELLS. In Figure 4 we labeled each region on the sphere by a vector $(\sigma_1, \sigma_2, \sigma_3)$ of signs $+, -$. In this section we give a similar encoding of the regions, as well as the vertices and edges, for an arbitrary arrangement. This encoding is used in the proof of our main theorem in Section 7. The sign vectors are easier to think about if we replace each cell e by the cone over e, i.e., by the set of positive scalar multiples of e. We begin by establishing the language for talking about these cones.

3.1. Chambers. The open regions into which $\mathbb{R}^3$ is cut by $\mathscr{A}$ are called *chambers*. For example, if $\mathscr{A}$ consists of the three coordinate hyperplanes, the chambers are the 8 open orthants. The regions on the sphere that we have been discussing are simply the intersections of the chambers with the sphere.

We can describe a chamber by specifying, for each $H \in \mathscr{A}$, which side of H the chamber is on. To formalize this, write $\mathscr{A} = \{H_i\}_{i \in I}$ and let H_i^+ and H_i^- be the two open halfspaces determined by H_i. (The choice of which one to call H_i^+ is arbitrary but fixed.) Then the chambers are precisely the nonempty sets of the form

$$C = \bigcap_{i \in I} H_i^{\sigma_i},$$

where $\sigma_i = \pm$. The vector of signs $(\sigma_i)_{i \in I}$ provides a succinct description of C. Figure 8 illustrates this (using the spherical picture) for the four planes $x = 0$, $y = 0$, $z = 0$, $z - x - y = 0$. The region labeled $++++$, for instance, corresponds to the chamber $x > 0$, $y > 0$, $z > 0$, $z - x - y > 0$. Notice that the sign vectors of the remaining chambers can be filled in once $++++$ has been identified; just move from chamber to chamber, crossing one plane at a time and recording the appropriate sign change.

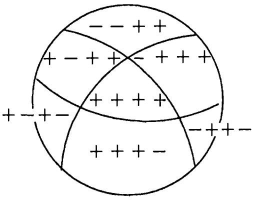


Figure 8. Sign vectors for four planes in general position.

3.2. Faces, products, and projections. The chambers are defined by finitely many linear inequalities, so they are polyhedra. They have faces, which are the nonempty sets obtained by changing zero or more inequalities to equalities. For example, the chamber $++++$ in Figure 8 has a face defined by $x = 0$, $y > 0$, $z > 0$, $z - x - y > 0$. We say that this face has sign vector $0+++$. In general, the sign vector of a face F is the vector (σ_i) such that

$$F = \bigcap_{i \in I} H^{\sigma_i},$$

where $\sigma_i \in \{+,-,0\}$ and $H_i^0 = H_i$. In Figure 8, the chamber $++++$ has three faces of codimension 1, with sign sequences $0+++$, $+0++$, and $+++0$. These correspond to the three sides of the triangle labeled $++++$ in Figure 8. Note that $++0+$ does not occur, because it would correspond to the inconsistent set of conditions $x > 0$, $y > 0$, $z = 0$, $z - x - y > 0$.

Let $\mathscr{C}$ be the set of chambers and let $\mathscr{F}$ be the set of faces of chambers; note that $\mathscr{C} \subset \mathscr{F}$ according to our conventions. Somewhat surprisingly, there is a naturally defined product on $\mathscr{F}$ that makes it a semigroup: Given $F, G \in \mathscr{F}$ with sign vectors $\sigma(F)$, $\sigma(G)$, their product FG is the face with sign vector

$$\sigma_i(FG) = \begin{cases} \sigma_i(F) & \text{if } \sigma_i(F) \neq 0 \\ \sigma_i(G) & \text{if } \sigma_i(F) = 0. \end{cases}$$

This has a geometric interpretation: If we move on a straight line from a point of F toward a point of G, then FG is the face we are in after moving a small positive distance.

Given $F \in \mathscr{F}$ and $C \in \mathscr{C}$, their product FC is again a chamber, called the *projection* of C on F. It can be characterized as the nearest chamber to C having F as a face. Here "nearest" refers to the number of hyperplanes in $\mathscr{A}$ separating C from FC. This proves the existence of the projection operators that were used to construct the random walk in Section 1.

We continue to identify chambers and faces with their intersections with the sphere. Thus we have a cell decomposition of the sphere, with a semigroup structure on the set of cells. The next subsection illustrates the use of sign vectors and the semigroup structure.

3.3. The diameter of the walk in $\mathbb{R}^3$. Consider an arrangement in $\mathbb{R}^3$ and the walk on regions defined in Section 1. It is easy to see that there is an integer $l > 0$ such that $K^l(C, C') > 0$ for all C, C'; in other words, one can get from any region to any other in l steps. The smallest such l is called the *diameter* of the walk. As an example, consider the arrangement of three planes shown (in its spherical representation) in Figure 1. By inspection, to get from a chamber to its antipode by steps of the walk (through a choice of vertices) takes three steps, and this is in fact the diameter.

Proposition 2. *The diameter of the walk based on n planes in* $\mathbb{R}^3$ *is always either* 2 *or* 3. *It is* 2 *if* $n \geq 5$ *and the planes are in general position.*

Proof: One can never get from a region C to its antipode $-C$ in one step; in fact, each vertex v has at least two 0's in its sign vector, so multiplying by v cannot change all the signs of C. This shows that the diameter is at least two.

Consider two regions C, C', with C' an m-gon. Assume, to simplify notation, that the sign vector σ' of C' has $\sigma_j' = +$ for all j and that the first m components

correspond to the m sides of C'. Then C' is the unique region whose sign vector has $+$ in its first m positions. If $m \geq 4$, we can find vertices v, w of C' such that v is on two of the sides of C' and w is on two different sides. Thus the sign vector of v has 0 in two of the first m slots and $+$ in the rest, and similarly for w, with the 0's of v and w in disjoint positions. The product vwC in the semigroup $\mathscr{F}$ therefore has $+$ in its first m positions, i.e., $vwC = C'$, and we can get from C to C' in two steps. This does not work if C' is a triangle ($m = 3$), but in that case we can get from any C to C' by using all three of the vertices of C'. Thus we have proved that the diameter is at most 3.

Finally, suppose that the planes are in general position, that $m = 3$, and that $n \geq 5$. Let w be a vertex on two of the great circles other than the three forming the sides of C'. In view of general position, the first 3 components of the sign vector of w are nonzero. Replacing w by $-w$ if necessary, we may assume that two of these 3 components, say the first two, are $+$. Let v be the vertex of C' whose sign vector has 0 in the first two slots. Then vw has $+$ everywhere, so we can get from C to C' in two steps. ■

4. HYPERPLANE ARRANGEMENTS. This section and the next are intended to provide some context for the seemingly strange random walk introduced in Section 1. We consider here arrangements of hyperplanes in spaces of arbitrary finite dimension d. The reader who wants to proceed to the proof of the main theorem may skip ahead to Section 6.

A good reference for the theory of hyperplane arrangements is the book by Orlik and Terao [**20**]. See also [**19**, **4**, **23**] and, for a concise summary of the basic concepts, [**6**, §2]. Throughout this section $\mathscr{A} = \{H_i\}_{i \in I}$ denotes a finite set of linear hyperplanes (subspaces of codimension 1) in a finite dimensional real vector space V. One often assumes that the intersection of the hyperplanes is the trivial subspace, as we have been assuming when $V = \mathbb{R}^3$. There is no loss of generality in making this assumption; for if it fails, then we can replace V by the quotient space V/V_0, where $V_0 = \bigcap_{i \in I} H_i$.

4.1. Chambers, faces, and products. The open regions into which V is cut by $\mathscr{A}$ are again called *chambers*. More precisely, the chambers are the connected components of the complement of $\bigcup_{i \in I} H_i$ in V. As in the 3-dimensional case, chambers and their faces are encoded by sign vectors $(\sigma_i)_{i \in I}$, where $\sigma_i \in \{+, -, 0\}$. The definition of the product of faces also remains unchanged, so we have a face semigroup $\mathscr{F}$ containing the set $\mathscr{C}$ of chambers, with $FC \in \mathscr{C}$ for $F \in \mathscr{F}$, $C \in \mathscr{C}$.

Remark. The faces are in 1–1 correspondence with their intersections with the unit sphere. If $\bigcap_{i \in I} H_i = \{0\}$, these intersections give a cell-decomposition of the sphere, as in the case $V = \mathbb{R}^3$.

4.2. A walk on the chambers. We may specify a random walk on $\mathscr{C}$ by assigning a weight w_F to each $F \in \mathscr{F}$. These weights satisfy $w_F \geq 0$ (many may be zero) and $\sum_{F \in \mathscr{F}} w_F = 1$. Given a starting chamber C_0, the walk proceeds by choosing from $\mathscr{F}$ with replacement, with w_F the chance of picking F each time. This generates choices $F_1, F_2, F_3, \ldots$. The walk proceeds by multiplying by F_i at stage i, i.e., by projecting onto F_i. Thus it goes

$$C_0, F_1 C_0, F_2 F_1 C_0, \ldots, F_i F_{i-1} \cdots F_1 C_0, \ldots$$

The chance of moving from C to C' in one step is

$$K(C,C') = \sum_{FC=C'} w_F. \tag{5}$$

The walk of Section 1 is based on n planes in $\mathbb{R}^3$ with $w_F = 0$ unless F is a half line (so that its intersection with the sphere is a point), and $w_F = 1/f_0$ for all half lines.

4.3. Analysis of the walk. As we remarked in the introduction, the hyperplane walks admit a rather complete theory. Bidigare, Hanlon, and Rockmore [**2**] give all the eigenvalues of the matrix K in a simple closed form. (We stated a special case of their result in Section 1.) Brown and Diaconis [**6**] prove diagonalizability of K and calculate eigenvalues using a chain complex given by the decomposition of the sphere by the hyperplanes. Brown [**5**] translates the walk to a semigroup setting and works out the appropriate Fourier analysis to get another proof of these results. The following result from [**6**] describes the stationary distribution and gives a bound on the convergence rate:

Theorem 2. *Let $\mathscr{A}$ be a hyperplane arrangement. Let $\{w_F\}$ be a probability distribution on the set $\mathscr{F}$ of faces. Let $K(C,C')$ be defined by* (5). *Then*

(a) *K has a unique stationary distribution π if and only if the weights $\{w_F\}$ are not concentrated on the faces in a single hyperplane, i.e., if and only if for each $H \in \mathscr{A}$ there is $F \not\subseteq H$ with $w_F > 0$.*
(b) *If the condition in* (a) *holds, then π may be described as follows*: *Sample* without *replacement from $\mathscr{F}$ according to the weights w_F. This generates an ordering $F_1, F_2, \dots, F_m$ of $\{F \in \mathscr{F}: w_F > 0\}$. The product $F_1F_2 \cdots F_mC_0$ is independent of C_0 and is a chamber distributed according to π.*
(c) *If the condition in (a) holds, then for any $C_0 \in \mathscr{C}$ and positive integer l*

$$\left\|K^l_{C_0} - \pi\right\| \le \sum_{H\in A} \lambda_H^l, \quad \textit{where } \lambda_H = \sum_{F\subseteq H} w_F.$$

Part (c) of Theorem 2 immediately gives the convergence rate bound stated in Section 1. And part (b) of Theorem 2 can be used to get the explicit formula for π in the 3-dimensional case, as stated in Theorem 1. Indeed, we first discovered the formula by doing exactly that. It turns out, however, that one can give an easier and completely self-contained proof of Theorem 1 by a direct argument based on equation (1). That is what we do in Section 7.

5. TWO EXAMPLES. There are many hyperplane arrangements where the chambers can be labeled in a natural way by familiar combinatorial objects such as permutations or trees and the walk described in Section 4.2 captures a natural mixing process. In this section we describe two such examples. The first is the braid arrangement, for which the chambers correspond to permutations. The action of faces on chambers gives natural shuffling schemes, such as the usual method of riffle shuffling a deck of cards, or a list rearrangement scheme used in computer science where a card is removed and replaced on top. The second example shows how hyperplane arrangements are related to tilings. In particular, we describe in some detail a hyperplane arrangement in $\mathbb{R}^3$ whose chambers correspond to rhombic tilings of a 10-gon. See [**2**, **6**] for further examples.

5.1. The braid arrangement and card shuffling. The *braid arrangement* in $\mathbb{R}^d$ consists of the $\binom{d}{2}$ hyperplanes $H_{ij} = \{(x_1, \dots, x_d) : x_i = x_j\}$ $(i < j)$. The chambers are associated with a common ordering of the coordinates and so with one of the $d!$ permutations. When $d = 4$, for example, one of the 24 chambers is the region defined by $x_1 > x_4 > x_2 > x_3$, corresponding to the permutation 1423.

The hyperplanes H_{ij} intersect in the line $x_1 = \cdots = x_d$. The braid arrangement therefore gives rise to an arrangement in a $(d-1)$-dimensional space, as explained at the beginning of Section 4. When $d = 4$, the resulting arrangement of 6 planes in $\mathbb{R}^3$ may be pictured as in Figure 9. The great circle corresponding to H_{ij} is labeled *i-j*. (The equator is *not* one of the great circles of the arrangement.) Each chamber is labeled with the associated permutation.

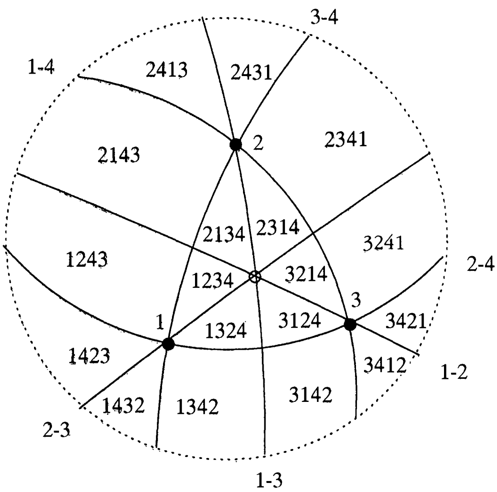


Figure 9. The braid arrangement when $d = 4$.

The faces of a chamber C are obtained by changing to equalities some of the inequalities defining C. For example, the chamber $x_2 > x_3 > x_1 > x_4$ has a face given by $x_2 > x_3 > x_1 = x_4$, which is also a face of the chamber $x_2 > x_3 > x_4 > x_1$. This common face is represented by the edge between 2314 and 2341 in Figure 9. Similarly, the vertex labeled 2 in the figure corresponds to the face $x_2 > x_1 = x_3 = x_4$; it is a face of six chambers, corresponding to the six possible orderings of the indices $1, 3, 4$.

It is useful to encode the system of equalities and inequalities defining a face F by an ordered partition $(B_1, \dots, B_k)$ of $\{1, \dots, d\}$. Here $B_1, \dots, B_k$ are disjoint nonempty sets whose union is $\{1, \dots, d\}$; they are called the *blocks* of the partition, and their order counts. For example, the face $x_2 > x_3 > x_1 = x_4$ corresponds to the 3-block ordered partition $(\{2\}, \{3\}, \{1, 4\})$, and the face $x_2 > x_1 = x_3 = x_4$ [vertex 2 in Figure 9] corresponds to the 2-block ordered partition $(\{2\}, \{1, 3, 4\})$. Notice that there is also a (unique) 1-block ordered partition, corresponding to the face $x_1 = x_2 = x_3 = x_4$. When we pass from $\mathbb{R}^4$ to a 3-dimensional quotient to make the hyperplanes have trivial intersection, this face becomes $\{0\}$. It does not show up in Figure 9 because its intersection with the sphere is empty.

The action of faces on chambers is easily pictured by thinking of a permutation $\tau = (\tau_1, \dots, \tau_d)$ as the set of labels on a deck of d cards, with the card labeled τ_1 on top, and so on. The ordered partition $B = (B_1, B_2, \dots)$ operates on τ by

removing cards with labels in B_1 and placing them on top (keeping them in the same relative order), then removing cards with labels in B_2 and placing them next, and so on. Suppose, for example, that $d = 10$, $\tau = (1, 7, 3, 9, 10, 4, 5, 2, 6, 8)$, and $B = (\{2, 5\}, \{3, 4, 6, 10\}, \{7\}, \{1, 8, 9\})$; then B acting on τ gives $(5, 2, 3, 10, 4, 6, 7, 1, 9, 8)$.

The walk on chambers described in Section 4 now yields a variety of shuffling schemes, depending on the choice of weights w_F. One example that has received much attention is the "Tsetlin library," or "random-to-top" shuffle. Here one assigns a positive weight w_i to each 2-block ordered partition $(\{i\}, \{1, \dots, d\} \setminus \{i\})$, and weight 0 to the other faces. When $d = 4$, for example, the vertices with positive weight are those labeled 1, 2, 3 in Figure 9 and the one that would be labeled 4 if it were visible. [The opposite vertex is shown as an open circle.] In the resulting random walk, a card is repeatedly picked at random according to the weights w_i and is replaced on top. This walk is useful in connection with self-organizing list-management schemes [**9**]. Imagine, for instance, a stack of d files, where file i is used with frequency w_i. Each time a file is used, it is replaced on top of the stack. After the process has been running for a long time, the most frequently used files will tend to be near the top.

The basic walk studied in this paper, where the vertices are weighted uniformly, is itself quite interesting when applied to the braid arrangement. Here, we assign weight $1/(2^d - 2)$ to each of the $2^d - 2$ two-block ordered partitions. The corresponding shuffling mechanism consists of "inverse riffle shuffles." In an ordinary riffle shuffle a deck of cards is divided into two piles, which are riffled together. The inverse chooses a set S of cards, which are removed ("unriffled") and placed on top; the $2^d - 2$ proper nonempty subsets S are all equally likely to be unriffled. This is very closely related to a standard model for riffle shuffling. See [**2**, **6**] for further details and references to earlier work.

To conclude, let us apply the results of Section 1.3 to the walk when $d = 4$ and the vertices are chosen uniformly. There are 14 vertices, 36 edges, and 24 regions. The stationary distribution is uniform because each of the regions has three sides. The eigenvalues and multiplicities (λ, m_λ) are $(1, 1)$, $(\frac{3}{7}, 6)$, $(\frac{1}{7}, 11)$, and $(0, 6)$. After shuffling l times, the distance to stationarity satisfies

$$\|K^l - \pi\| \le 6\left(\frac{3}{7}\right)^l.$$

This bound shows that $l = 5$ shuffles suffice to make the distance to stationarity smaller than $1/10$.

5.2. Arrangements and tilings. We discuss now an example of an arrangement in $\mathbb{R}^3$ in which the chambers correspond to certain tilings of a centrally symmetric 10-gon. The walks described in this paper can then be used to generate random tilings. Such tilings are of interest to physicists who study quasicrystals; see [**8**, **16**] and the references cited there. This is an example of a very general theory whereby hyperplane arrangements can be associated with a set of tilings of special convex polytopes called zonotopes [**3**, §4]. A good introductory reference for the material in this subsection is [**23**, Lecture 7]

Given n pairwise linearly independent vectors $v_1, \dots, v_n \in \mathbb{R}^d$, we define a convex polytope known as a *zonotope* by

$$Z(v_1, \dots, v_n) := \left\{ \sum_i \lambda_i v_i | 0 \le \lambda_i \le 1 \right\}.$$

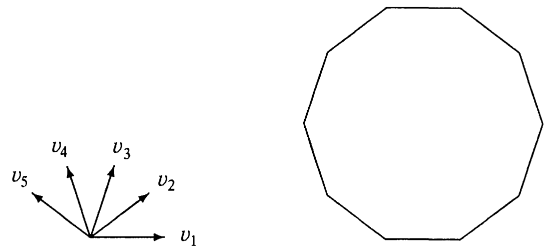


Figure 10. Five vectors in $\mathbb{R}^2$ and the 10-gon they generate.

We call the v_i the *zones* of $Z(v_1, \ldots, v_n)$. When $d = 2$, $Z(v_1, \ldots, v_n)$ is a centrally symmetric $2n$-gon; this is illustrated in Figure 10 for the case $d = 2$ and $n = 5$, where the resulting zonotope is a 10-gon.

We are interested in tilings of this 10-gon by the parallelograms generated by the $\binom{5}{2} = 10$ pairs v_i, v_j, $1 \le i < j \le 5$. Since we may take all the v_i to have the same length without loss of generality, we call these tilings *rhombic* tilings. Two such rhombic tilings are illustrated in Figure 11. It is a consequence of the work of Billera and Sturmfels [3] that the set of all such tilings of this 10-gon corresponds to the chambers of an arrangement of $\binom{5}{3} = 10$ hyperplanes in $\mathbb{R}^3$. We describe this arrangement in some detail now.

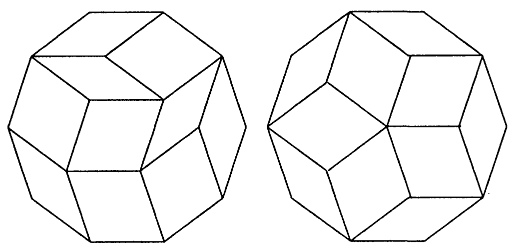


Figure 11. Two tilings of the 10-gon by parallelograms.

Every triple v_i, v_j, v_k of the 5 given vectors, $1 \le i < j < k \le 5$, satisfies a unique (up to nonzero scalar) linear dependence relation. We write it as a relation among v_1, v_2, v_3, v_4, v_5, so it is given by a vector of length 5. Let $\tau = \{i, j, k\}$ and let the dependence be given by $z^\tau \in \mathbb{R}^5$; thus $\sum_{l=1}^5 z_l^\tau v_l = 0$ and $z_l^\tau = 0$ if $l \notin \tau$. The span V of the z^τ, over all triples $\tau \subset \{1, 2, 3, 4, 5\}$, is the nullspace of the rank 2 matrix

$$A = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ y_1 & y_2 & y_3 & y_4 & y_5 \end{pmatrix}.$$

where $v_i = (x_i, y_i)$, and so is 3-dimensional. For each triple τ, we denote by $H^\tau \subset \mathbb{R}^5$ the hyperplane having z^τ as its normal (i.e., $H^\tau = \{x \in \mathbb{R}^5 | \langle x, z^\tau \rangle = 0\}$). The hyperplanes $\{H^\tau\}$ have a 2-dimensional intersection $V^\perp$ (the row space of A). We therefore get an arrangement in the 3-dimensional space $\mathbb{R}^5/V^\perp$ (which we may identify with V), as explained at the beginning of Section 4. This may be pictured as in Figure 12. Each great circle is labeled by the corresponding triple τ. Note that the 4 great circles corresponding to triples $\tau \subset \{2, 3, 4, 5\}$ intersect at a pair of antipodal vertices, one of which is labeled 1 in Figure 12. This can be

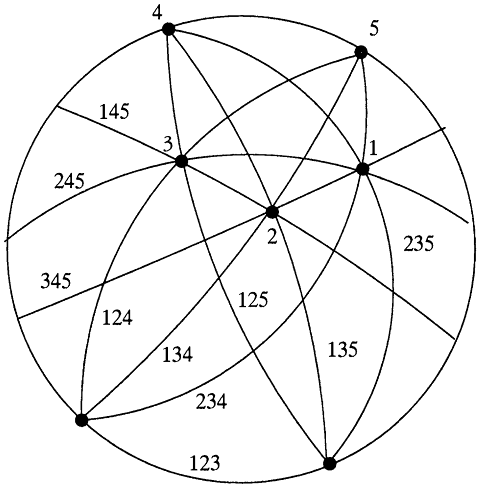


Figure 12. The arrangement associated to tilings of a 10-gon.

explained by straightforward linear algebra, which we leave to the interested reader. [Hint: Consider the matrix obtained from A by deleting the first column.] The other multiple intersections (with labels $2, \dots, 5$) can be explained in a similar way.

We claim that the chambers of this arrangement correspond to the rhombic tilings of the 10-gon. To understand why this might be so, take any point $\theta = (\theta_1, \dots, \theta_5) \in \mathbb{R}^5$ not on any of the hyperplanes H^τ. Lift the zones v_i to $\tilde{v}_i = (x_i, y_i, \theta_i)$, $i = 1, \dots, 5$, and form the 3-dimensional zonotope $\tilde{Z} := Z(\tilde{v}_1, \dots, \tilde{v}_5) \subset \mathbb{R}^3$. One can show that each 2-dimensional face of $\tilde{Z}$ is a parallelogram. [Brief explanation: One first observes that the 2-dimensional faces of a 3-dimensional zonotope are translates of planar zonotopes spanned by subsets of the generating zones; see [**23**, p. 205]. These are all parallelograms unless three of the lifted zones $\tilde{v}_i, \tilde{v}_j, \tilde{v}_k$ are linearly dependent. A linear dependence relation among these lifts would be a multiple of z^τ, where $\tau = \{i, j, k\}$. This would imply $\theta \in H^\tau$, contradicting the choice of θ.] Projecting the faces on the bottom of $\tilde{Z}$ (i.e., the faces whose outer normal has a negative third coordinate) onto the 10-gon $Z(v_1, \dots, v_5)$, one gets a rhombic tiling of the 10-gon. Further argument shows that any other θ in the same chamber gives the same tiling and that all rhombic tilings of the 10-gon arise in this way.

An example of what might happen if the original zonotope were a hexagon and the lifted zonotope a 3-cube is shown in Figure 13; depending on its orientation in $\mathbb{R}^3$, the bottom of the cube would look like one of two tilings shown.

A move between adjacent chambers in the arrangement in Figure 12 corresponds to a simple local change in the corresponding tiling: for some embedded "3-cube" in the tiling, flip just that portion of the tiling as shown in Figure 13. So

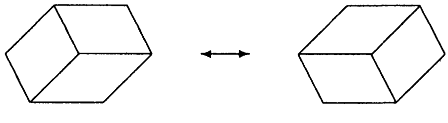


Figure 13. A flip of a 3-cube.

the chamber corresponding to a tiling is an i-gon if and only if the tiling has i embedded 3-cubes; for example the tiling on the left in Figure 11 corresponds to a triangular region, while that on the right corresponds to a pentagonal region.

We note that the situation described here is a piece of a very general theory of subdivisions [**3**, **1**]. In particular, given a d-dimensional zonotope Z with n zones in $\mathbb{R}^d$, one can associate with it an arrangement of at most $\binom{n}{d+1}$ hyperplanes in $\mathbb{R}^{n-d}$ whose chambers correspond to certain tilings of Z. In our example it turns out to be *all* the rhombic tilings, but, in general, not all tilings can be obtained by the lifting procedure described in the example; see [**3**] for details. When $n = d + 3$ the derived arrangement is 3-dimensional, and so theory of the present paper applies.

To conclude, let us apply the results of Section 1.3 to our walk on the tilings of a 10-gon. There are 40 vertices, 100 edges, and 62 regions, with $(p_3, p_4, p_5) = (50, 10, 2)$. The stationary probability of a tiling with i cubes is $(i-2)/76$. The eigenvalues and multiplicities (λ, m_λ) are $(1,1)$, $(\frac{1}{4}, 10)$, $(\frac{1}{20}, 30)$, and $(0, 21)$. After l steps, the distance to stationarity satisfies

$$\|K^l - \pi\| \leq 10\left(\frac{1}{4}\right)^l.$$

6. A GEOMETRIC DESCRIPTION OF K. We return to the setting of Section 1, where $\mathscr{A}$ is a collection of planes in $\mathbb{R}^3$ with trivial intersection, and K is the transition matrix of the walk on the regions of the sphere (or, equivalently, on the chambers in $\mathbb{R}^3$). Recall that, in the notation of Section 1,

$$K(C, C') = \frac{1}{f_0}\Lambda(C, C'), \tag{6}$$

where $\Lambda(C, C')$ is the number of vertices v of C' such that $vC = C'$. Note that $\Lambda(C, C)$ is the number of vertices of C and that $\Lambda(C, -C) = 0$. The main result of this section is the following formula for $\Lambda(C, C')$ when $C' \neq \pm C$: If k is the number of sides of C' for which the sign vectors of C and C' agree (that is, the number of bounding planes of C that do *not* separate C from C'), then

$$\Lambda(C, C') = k - 1.$$

We prove this in Section 6.2 (Lemma 2). We begin by describing the matrix Λ in terms of lunes.

6.1. Lunes. For a region D and a vertex v of D, let $l(v, D)$ be the *lune* determined by v and D as in Figure 14. Thus if H_1 and H_2 are the great circles bounding D and passing through v, and if H_i^+ is the hemisphere defined by H_i

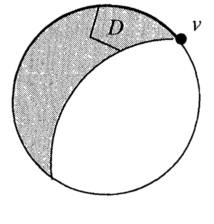


Figure 14. The lune $l(v, D)$.

that contains D $(i = 1, 2)$, then $l(v, D)$ is the intersection $H_1^+ \cap H_2^+$. Given two regions C, C', we define the *lunacy* of C with respect to C' to be the number of vertices v of C' such that $C \subseteq l(v, C')$.

Proposition 3. *For any two regions C, C', $\Lambda(C, C')$ is equal to the lunacy of C with respect to C'.*

Proof: Assume, to simplify notation, that the sign vector of C' consists of all $+$'s. For a vertex v of C', we have $vC = C'$ if and only if the sign vector of vC is $+$ in the positions corresponding to the sides of C'. Now v has $+$ in all of these positions except two, where it has 0; so $vC = C'$ if and only if C already has $+$ in the two positions corresponding to the sides of C' that intersect at v. But this says precisely that $C \subseteq l(v, C')$. ■

Understanding K, then, reduces to calculating lunacy. Figure 15 shows a simple example; here C is in the lune $l(v, C')$ but not the other two C'-lunes, so $\Lambda(C, C') = 1$. On the other hand, C' is in two of the four C-lunes, so $\Lambda(C', C) = 2$. Note the lack of symmetry.

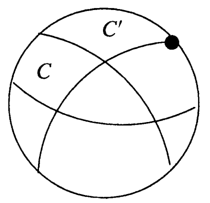


Figure 15. $\Lambda(C, C') = 1$.

There is a similar description of the transition matrix for the higher-dimensional hyperplane walks described in Section 4. One replaces the lunes by sets $l(F, C)$, where C is a chamber, F is a face of C, and $l(F, C)$ is an intersection of open half-spaces, one for each bounding hyperplane of C. What's special about the 3-dimensional case, however, is that lunacy is easy to compute. The computation depends on a simple geometric fact about polygons, related to "shellability."

6.2. Shellability and the calculation of lunacy. Consider a planar convex polygon P whose sides are extended to form lines $l_1, \ldots, l_n$. If the polygon is viewed from any point x outside of it, one sees a contiguous set of its edges, and these are precisely the edges whose lines separate x from P. See Figure 16; here it is

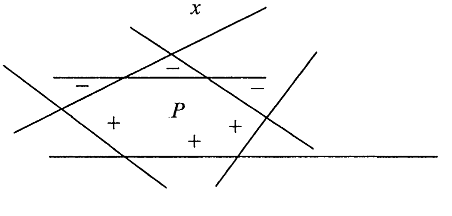


Figure 16. The sides visible from x.

assumed that P is on the positive side of each l_i, and the lines are labeled with $\pm$ to indicate which side x is on.

All of this works equally well on the sphere:

Lemma 1 (Shelling lemma). *Given a spherical polygon bounded, in cyclic order, by n great circles $l_1, \dots, l_n$, suppose that the interior of the polygon is on the positive side of each of these circles. Then any point on the sphere not on any of the circles is on the positive side of a cyclically contiguous subset of them.*

An analogue of the shelling lemma holds for convex polytopes in d-dimensional space [7] and for oriented matroids [18]. In the former case, contiguity is replaced by "shellability" of the set of visible faces, whence the name "shelling lemma." The latter case implies that everything we have done works for arrangements of pseudocircles (Section 2.4). See [23] and [4] for discussions of all of these topics. We omit a proof of the shelling lemma because it is treated thoroughly in the references cited and because, in the present low-dimensional case, it is quite believable from the picture.

Lemma 2 (Lunacy lemma). *Given two regions C and C' that are neither equal nor opposite, we have*

$$\Lambda(C, C') = k - 1,$$

where k is the number of sides of C' for which the sign vectors of C and C' agree. Consequently,

$$\Lambda(C, C') + \Lambda(-C, C') = i - 2$$

if C' is an i-gon.

In Figure 15, for example, C' has 3 sides, one of which separates C from C'; so $k = 2$ and $\Lambda(C, C') = 1$, as we have already observed.

Proof of the lunacy lemma. Assume that C' has sign vector $+ + \cdots +$ and apply the shelling lemma to the polygon C', with x equal to any interior point of C. It follows that the sign pattern of C, restricted to the sides of C', is a cyclic permutation of the string $+ + \cdots + - - \cdots -$, where there are k $+$'s. Every pair of cyclically consecutive $+$'s corresponds to a lune of C' that contains C. Since there are $k - 1$ such pairs, the lunacy $\Lambda(C, C')$ is indeed $k - 1$. To prove the second assertion of the lemma, note that for each of the i sides of C' either C or $-C$ has the same sign as C'. Thus the k's for C and $-C$ sum to i. ∎

Remark. If $C = C'$, then $\Lambda(C, C') = i$, while if $C = -C'$ then $\Lambda(C, C') = 0$. Hence $\Lambda(C, C') + \Lambda(-C, C') = i$ rather than $i - 2$ if $C = \pm C'$.

6.3. Convergence rate examples. Recall from Section 2.1 that the walk on regions is close to stationary after 2 steps for an arrangement of n planes in general position if n is large. Now that we understand the matrix K, we can give an example where the walk is not close to stationary after 1 step, as advertized in Section 2.1. In other words, we want $\|K_C - \pi\|$ to be bounded away from 0 for an infinite family of general position arrangements with larger and larger n. We use a family of arrangements $\mathscr{A}_n$ constructed by Füredi–Palásti [10], for which p_3, the number of triangles, is asymptotic to $2n^2/3$ as $n \to \infty$. (Recall from Section 2.1 that f_2, the total number of regions, is about n^2. So there are about $n^2/3$

non-triangular regions. Most of these are hexagons.) Fix a starting region C. We claim that for any triangular region C', either $\Lambda(C, C') = 0$ or $\Lambda(C, -C') = 0$. To see this, we may assume $C' \neq \pm C$. Then $\Lambda(C, C') + \Lambda(-C, C') = 1$ by the lunacy lemma, so one term is 0. Our claim now follows from the fact that $\Lambda(-C, C') = \Lambda(C, -C')$.

In view of the claim, there are $p_3/2$ triangular regions C' that cannot be reached from C in one step. Our main theorem, however, implies that this set $\mathscr{T}$ of triangles has stationary probability

$$\pi(\mathscr{T}) = \frac{p_3}{2} \cdot \frac{1}{2(f_0 - 2)}.$$

The definition (2) of the distance between probability distributions now implies that $\|K_C - \pi\| \geq \pi(\mathscr{T})$, which tends to $1/6$ as $n \to \infty$. Thus the walk for $\mathscr{A}_n$ is not close to stationary after one step.

Remark. On the other hand, there is a naturally occurring family of arrangements $\mathscr{B}_n$ for which the walk *is* close to random after one step. Here $\mathscr{B}_n$ is the *cyclic* arrangement, consisting of n great circles that form the sides of an n-gon in the Northern Hemisphere. There is, of course, also an n-gon antipodal to this one in the Southern Hemisphere. $\mathscr{B}_n$ is a general position arrangement; for $n \geq 5$ there are $2n$ triangles and $n(n-3)$ quadrilaterals in addition to the two n-gons. The cases $n = 3, 4, 5, 6$ are shown in Figures 1, 2, 5, and 6. A straightforward but tedious calculation shows that, for any starting region C, $\|K_C - \pi\| \leq c/n$ for c a universal constant; thus the walk for this family *is* close to random after one step when n is large.

7. PROOF OF THEOREM 1. We are trying to calculate the stationary distribution of the walk with transition matrix given by (6). The assertion is that

$$\pi(C) = \frac{i-2}{2(f_0 - 2)} \tag{7}$$

if C is an i-gon. There are two ways to proceed.

Method 1. As we remarked in the introduction, it suffices to show that the right-hand side of (7) is a probability distribution satisfying

$$\sum_C \pi(C) K(C, C') = \pi(C') \tag{8}$$

for all C'. Now we already know, by the Euler relation, that the right-hand side sums to 1 (Section 2.2). So it suffices to show, for each C', that

$$\sum_C (i(C) - 2) K(C, C') = i(C') - 2,$$

where $i(C)$ is the number of sides of C. Equivalently, we must show

$$\sum_C (i(C) - 2) \Lambda(C, C') = f_0(i(C') - 2). \tag{9}$$

Since $i(-C) = i(C)$, we can replace $\Lambda(C, C')$ on the left by the average

$$M(C, C') = \frac{\Lambda(C, C') + \Lambda(-C, C')}{2}$$

without changing the value of the sum. In view of the lunacy lemma, the left side of (9) becomes

$$\sum_C (i(C) - 2)M(C,C') = \sum_C (i(C) - 2)\frac{i(C') - 2}{2} + 2(i(C') - 2),$$

where the second term takes account of the fact that $M(\pm C', C') = i(C')/2$ instead of $(i(C') - 2)/2$. Pulling out the factor $i(C') - 2$, and recalling that $\Sigma_C(i(C) - 2) = 2(f_0 - 2)$, we obtain (9).

Method 2. We know that π exists and is characterized by (8), which can be rewritten

$$\sum_C \pi(C)\Lambda(C,C') = f_0\pi(C'). \tag{10}$$

Since $\Lambda(C,C') = \Lambda(-C, -C')$, it follows that $\pi(-C) = \pi(C)$. We may therefore rewrite (10) as

$$\sum_C \pi(C)M(C,C') = f_0\pi(C'),$$

with M as in the first proof. Using the lunacy lemma (and taking account of the terms in the sum where $C = \pm C'$), the equation becomes

$$\sum_C \pi(C)\frac{i(C') - 2}{2} + 2\pi(C') = f_0\pi(C')$$

or, since $\Sigma_C \pi(C) = 1$,

$$\frac{i(C') - 2}{2} + 2\pi(C') = f_0\pi(C').$$

Solving for $\pi(C')$, we obtain

$$\pi(C') = \frac{i(C') - 2}{2(f_0 - 2)},$$

as required. ∎

This second proof is somewhat more satisfying than the first, since equation (7) appears naturally, rather than being pulled out of a hat. Note also that the second proof did not use the Euler relation; it therefore has the Euler relation as a consequence. Note, finally, that Method 2 proves uniqueness of π (though not existence) without appeal to the theory of Markov chains.

Remark. We are mystified by the factor $i - 2$ in Theorem 1. To help appreciate the mystery (and Theorem 1) let us consider a different random walk on the chambers of a hyperplane arrangement, called the *local walk*. Suppose the walk is currently in region C bounded by i hyperplanes. Pick one of the bounding planes uniformly at random. The walk moves to the chamber adjacent to C along the chosen hyperplane. Call the transition matrix of this walk $H(C,C')$. If $i(C)$

denotes the number of sides of C, then

$$H(C,C') = \begin{cases} 1/i(C) & \text{if } C' \text{ is adjacent to } C \\ 0 & \text{otherwise.} \end{cases}$$

For this walk, an i-sided chamber C has stationary probability proportional to $i = i(C)$; moreover, this walk is reversible: $i(C)H(C,C') = i(C')H(C',C)$. The more vigorous walk $K(C,C')$ that we have been studying is not reversible except in special cases. It is one of only a few families of nonreversible walks where the stationary distribution is known explicitly. For the braid arrangement of Section 5.1, the local walk corresponds to mixing a deck of cards by adjacent transpositions, whereas the walk of this paper corresponds to the (much faster) riffle shuffling. For tilings (Section 5.2), physicists currently use the local walk (cube flips) for simulating a random tiling; see **[22]**. The walk described here may offer vast speed-ups.

8. FINAL REMARKS. The plane arrangements $\mathscr{A}$ in $\mathbb{R}^3$ that we have been considering in this paper occur in the literature under the name *line arrangements*. In fact, a plane through the origin in $\mathbb{R}^3$ is the same as a line in the projective plane P^2, so we can think of $\mathscr{A}$ as an arrangement of lines in the plane. Recall that the points of P^2 are the lines through the origin in $\mathbb{R}^3$. Since a line through the origin in $\mathbb{R}^3$ corresponds to a pair of antipodal points on the sphere S^2, we get a 2-to-1 map $S^2 \to P^2$. The great circles in our spherical pictures are the inverse images of the lines in P^2 under this map. As a practical matter, if P^2 is viewed as the affine plane with an extra line at infinity, one goes from the projective picture to the spherical picture by drawing the projective picture on the Northern Hemisphere, the line at infinity becoming the equator.

For a simple example, consider the arrangement of four lines in P^2 consisting of the three lines shown in Figure 17 together with the line at infinity. The

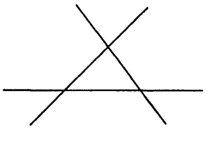


Figure 17. Four lines in P^2; the line at infinity is included.

corresponding spherical picture is Figure 2. A more complicated example of a projective picture was given in Figure 7. The cell-decomposition of S^2 that we have been using in this paper is obtained by lifting an analogous cell-decomposition of P^2. Thus there is a pair of antipodal cells in S^2 for each cell in P^2.

There is an extensive literature on line arrangements (and "pseudoline" arrangements) going back to the 1820's. Much is known, but there are still many open questions; see **[14, 11]** for surveys. Some of the more interesting results and open questions concern the polygon counts p_i ($i \geq 3$) where, following standard conventions, we now denote by p_i the number of i-gons in the cell decomposition of P^2; this is half of what was called p_i in Section 2. A sample result is that for an arrangement in general position, one has

$$p_3 = 4 + \sum_{i \geq 5} (i-4)p_i. \tag{11}$$

(The reader can check that this equation holds for each column of Table 2, after the numbers in that table are cut in half.) Moreover, given any sequence $p_3, p_5, p_6, p_7, \ldots$ of natural numbers satisfying (11), one can always find a general position arrangement with those polygon counts. One has no control over p_4, however, and, indeed, very little is known about p_4. Equivalently, one has no control over the number of lines in the arrangement. Thus we are still very far from knowing, for a given n, what vectors $(p_3, p_4, \ldots, p_n)$ can occur.

See [**14, 11**] for a wealth of further results, conjectures, and open questions.

ACKNOWLEDGMENTS. We thank Catherine Stenson and Eli Goodman for several helpful comments and references.

REFERENCES

1. Margaret M. Bayer and Keith A. Brandt, Discriminantal arrangements, fiber polytopes and formality, *J. Algebraic Combin.* **6** (1997) 229–246.
2. T. Patrick Bidigare, Phil Hanlon, and Daniel N. Rockmore, A combinatorial description of the spectrum for the Tsetlin library and its generalization to hyperplane arrangements, *Duke Math. J.*, to appear.
3. Louis J. Billera and Bernd Sturmfels, Fiber polytopes, *Ann. of Math.* (2) **135** (1992) 527–549.
4. Anders Björner, Michel Las Vergnas, Bernd Sturmfels, Neil White, and Günter M. Ziegler, *Oriented matroids*, Encyclopedia of Mathematics and its Applications, vol. 46, Cambridge University Press, Cambridge, 1993.
5. Kenneth S. Brown, Semigroups, rings, and random walks, preprint, 1998.
6. Kenneth S. Brown and Persi Diaconis, Random walks and hyperplane arrangements, *Ann. Probab.*, to appear.
7. H. Bruggesser and P. Mani, Shellable decompositions of cells and spheres, *Math. Scand.* **29** (1971) 197–205 (1972).
8. Veit Elser, Random tiling structure of icosahedral quasicrystals, *Philosophical Magazine B* **73** (1996) 641–656.
9. James Allen Fill, An exact formula for the move-to-front rule for self-organizing lists, *J. Theoret. Probab.* **9** (1996) 113–160.
10. Z. Füredi and I. Palásti, Arrangements of lines with a large number of triangles, *Proc. Amer. Math. Soc.* **92** (1984) 561–566.
11. Jacob E. Goodman, Pseudoline arrangements, in *Handbook of discrete and computational geometry* (J. E. Goodman and J. O'Rourke, eds.), CRC Press, Boca Raton, 1997, pp. 83–109.
12. Branko Grünbaum, *Convex polytopes*, Pure and Applied Mathematics, vol. 16, Wiley-Interscience, New York, 1967, with the cooperation of Victor Klee, M. A. Perles and G. C. Shephard.
13. ______, Arrangements of hyperplanes, in *Proceedings of the Second Louisiana Conference on Combinatorics, Graph Theory and Computing* (Louisiana State University, Baton Rouge, La., March 8–11, 1971) (R. C. Mullin, K. B. Reid, D. P. Roselle, and R. S. D. Thomas, eds.), Louisiana State University, 1971, pp. 41–106.
14. ______, *Arrangements and spreads*, CBMS Regional Conference Series in Mathematics, vol. 10, American Mathematical Society, Providence, R.I., 1972.
15. Branko Grünbaum and G. C. Shephard, Simplicial arrangements in projective 3-space, *Mitt. Math. Sem. Giessen* (1984), no. 166, 49–101.
16. Christopher L. Henley, Random tiling models, in *Quasicrystals: the state of the art* (D. P. DiVincenzo and P. J. Steinhardt, eds.), Directions in Condensed Matter Physics, vol. 11, World Scientific Publishing Co. Inc., River Edge, NJ, 1991, pp. 429–524.
17. John G. Kemeny and J. Laurie Snell, *Finite Markov chains*, The University Series in Undergraduate Mathematics, D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto-London-New York, 1960.
18. Arnaldo Mandel, *Topology of oriented matroids*, Ph.D. thesis, University of Waterloo, 1982.
19. Peter Orlik, *Introduction to arrangements*, CBMS Regional Conference Series in Mathematics, vol. 72, American Mathematical Society, Providence, R.I., 1989.
20. Peter Orlik and Hiroaki Terao, *Arrangements of hyperplanes*, Grundlehren der Mathematischen Wissenschaften, vol. 300, Springer-Verlag, Berlin, 1992.

21. Richard P. Stanley, *Enumerative combinatorics. Vol.* 1, Cambridge Studies in Advanced Mathematics, vol. 49, Cambridge University Press, Cambridge, 1997, with a foreword by Gian-Carlo Rota, corrected reprint of the 1986 original.
22. K. J. Strandburg, Lei-Han Tang, and M. V. Jaric, Phason elasticity in entropic quasicrystals, *Physical Review Letters* **63** (1989) 314–317.
23. Günter M. Ziegler, *Lectures on polytopes*, Graduate Texts in Mathematics, vol. 152, Springer-Verlag, New York, 1995.

LOUIS BILLERA studied mathematics at Rensselaer Polytechnic Institute and the City University of New York Graduate Center, with a year in between studying psychology at Princeton. He has been a faculty member at Cornell since 1968, in one or both of Operations Research and Mathematics. His early work was mostly in game theory and mathematical economics, but he has worked almost exclusively on questions relating convexity and combinatorics for the past 20 years. In 1994 he was awarded the Fulkerson Prize in Discrete Mathematics by the American Mathematical Society and the Mathematical Programming Society.
Cornell University, Ithaca, NY 14853-7901
billera@math.cornell.edu

KEN BROWN was an undergraduate at Stanford and a graduate student at M.I.T., where he studied with Dan Quillen. He has been at Cornell since 1971. Most of his research has involved topological methods in group theory, but his interests have broadened in recent years to include probability theory. He is the author of two books, *Cohomology of Groups* and *Buildings*.
Cornell University, Ithaca, NY 14853-7901
kbrown@math.cornell.edu

PERSI DIACONIS left high school at an early age to earn a living as a magician and gambler, only later to become interested in mathematics and earn a Ph.D. at Harvard. After a spell at Bell Labs, he has been a professor at Stanford, Harvard, and Cornell. He was an early recipient of a MacArthur Foundation award, and his wide range of mathematical interests is partly reflected in his first book, *Group Representations in Probability and Statistics*. He retains an interest in magic and the exposure of fraudulent psychics.
Stanford University, Stanford, CA 94305

A Counting Formula for Primitive Tetrahedra in $\mathbf{Z}^3$

Mizan R. Khan

1. INTRODUCTION. A *primitive polytope* is a polytope in n-dimensional Euclidean space $\mathbf{R}^n$ whose vertices are (integer) lattice points (points of $\mathbf{Z}^n$) but it does not contain any other lattice points in its interior or on its boundary. Pick's theorem coupled with the fact that any primitive (convex) polygon has at most 4 vertices (by Theorem 2.1) allows us to infer that a convex polygon in $\mathbf{Z}^2$ is primitive if and only if it is a triangle of area $\frac{1}{2}$ or a parallelogram of area 1. In $\mathbf{Z}^3$ matters are very different. Specifically, there is no bound on the volume of a primitive tetrahedron (see Section 3); but there is an elegant characterization of primitive tetrahedra that was discovered more than 30 years ago. This characterization gives us a simple formula that counts the number of equivalence classes of primitive tetrahedra of a given volume. The goal of this paper is to state and prove this formula.

We begin by defining the terms *lattice polytope*, *vertex*, and *primitive*.

Definition 1.1. A *lattice polytope* is the closed convex hull in $\mathbf{R}^n$ of a finite set $\{u_1, \dots, u_m\}$ of points of $\mathbf{Z}^n$, i.e., it is the set

$$\{t_1u_1 + \cdots + t_mu_m \colon u_1, \dots, u_m \in \mathbf{Z}^n, 0 \le t_1, \dots, t_m \le 1, t_1 + \cdots + t_m = 1\}.$$

A *vertex* of a polytope is a point of the polytope that cannot be written as a convex combination of other points of the polytope.

Remark. The lattice polytopes we consider always have *non-zero volume*.

Definition 1.2. A lattice polytope in $\mathbf{R}^n$ is *primitive* if the only lattice points it contains are its vertices.

Reeve ([**8**] and [**9**]) and Reznick [**10**] use the term *fundamental* instead of primitive; Scarf [**13**] uses the term *integral polyhedron*. Our preference for using the adjective *primitive* instead of *fundamental* or *integral* stems from the usage in [**14**]. The interested reader should look at [**14**, pp. 98–99] for a discussion of primitive sets of vectors and the connection with constructing a reduced basis of a lattice.

Section 2 is a brief discussion of primitivity in $\mathbf{Z}^n$. We explain why a primitive polytope in $\mathbf{Z}^n$ has no more than 2^n vertices and why a primitive parallelepiped in $\mathbf{Z}^n$ has unit volume. Section 3 is a comparison of primitive triangles in $\mathbf{Z}^2$ with primitive tetrahedra in $\mathbf{Z}^3$. The goals of this section are to explain why there is no bound on the volume of primitive tetrahedra and to present a geometric characterization of primitive tetrahedra. In Section 4 we discuss unimodular maps and state an analytic version of the geometric characterization. Section 5 is where we discuss a formula that counts the number of equivalence classes of primitive tetrahedra for a fixed volume; the equivalence relation is defined via unimodular maps. The

formula is in Theorem 5.3 and is proved with the aid of Burnside's lemma. The final section describes Howe's generalization of Theorem 4.2, which describes all primitive polyhedra, not just primitive tetrahedra. Except for the results in Section 2, we confine our remarks to dimension 3 because the characterizations of primitive tetrahedra do not extend to higher dimensions.

To conserve space we typically write lattice points as row vectors, but in our matrix calculations we view lattice points as column vectors.

2. PRIMITIVITY IN $\mathbf{Z}^n$.

We begin by showing that the number of vertices of a primitive polytope is bounded, a result that appears as Theorem 1.2 in [**13**].

Theorem 2.1. *A primitive polytope in* $\mathbf{R}^n$ *has at most* 2^n *vertices.*

Proof: If there are more than 2^n vertices in a lattice polytope P, then there are two vertices $v = (v_1, v_2, \ldots, v_n)$, $w = (w_1, w_2, \ldots, w_n)$ such that $v_i \equiv w_i \pmod 2$, $i = 1, \ldots, n$. The lattice point $(v + w)/2$ belongs to P, so P is not primitive. ∎

We now discuss lattice *parallelepipeds* in $\mathbf{R}^n$. We like to think of the lattice points in a lattice parallelepiped as the cosets of a quotient group. Suppose P is the parallelepiped $P = \{u + t_1v_1 + t_2v_2 + \cdots + t_nv_n\colon\ u, v_1, v_2, \ldots, v_n \in \mathbf{Z}^n,\ 0 \leq t_1, t_2, \ldots, t_n \leq 1\}$. We can consider the sublattice $(\mathbf{Z}v_1 \oplus \mathbf{Z}v_2 \oplus \cdots \oplus \mathbf{Z}v_n)$ and view the lattice points in P as the cosets of the quotient group $\mathbf{Z}^n/(\mathbf{Z}v_1 \oplus \mathbf{Z}v_2 \oplus \cdots \oplus \mathbf{Z}v_n)$. We have found the following result very useful.

Theorem 2.2. *Let* $v_1, \ldots, v_n$ *be a set of linearly independent points of* $\mathbf{Z}^n$. *Then the quotient group* $\mathbf{Z}^n/(\mathbf{Z}v_1 \oplus \mathbf{Z}v_2 \oplus \cdots \oplus \mathbf{Z}v_n)$ *has order* $|\det A|$, *where* A *is the* $n \times n$ *matrix whose k-th column is* v_k, *i.e., the order equals the volume of the parallelepiped spanned by* $v_1, v_2, \ldots, v_n$.

For a proof see [**14**, Theorem 20, p. 49]. Let us illustrate Theorem 2.2 and the remarks preceding it by examining the parallelogram P in $\mathbf{Z}^2$ whose vertices are $(0,0)$, $(1,3)$, $(2,2)$, and (3.5).

The area of P equals 4. Therefore the quotient group $\mathbf{Z}^2/(\mathbf{Z}(1,3) \oplus \mathbf{Z}(2,2))$ has order 4.

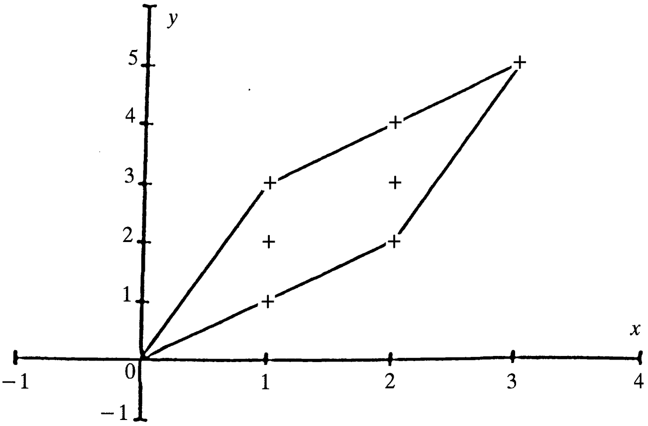


Figure 1. The parallelogram with vertices $(0,0)$, $(1,3)$, $(2,2)$, and $(3,5)$.

The 8 lattice points in P represent the cosets of $\mathbf{Z}^2/(\mathbf{Z}(1,3) \oplus \mathbf{Z}(2,2))$ in the following way: the zero coset can be represented by any one of the 4 vertices; the lattice points $(1,1)$ and $(2,4)$ represent the same coset (this coset is its own inverse); the lattice points $(1,2)$ and $(2,3)$ represent distinct cosets with one being the inverse of the other. Clearly, the quotient group is isomorphic to $\mathbf{Z}/4\mathbf{Z}$.

From Theorem 2.2 and our observation about viewing lattice points as cosets, we can easily prove the following theorem about parallelepipeds in $\mathbf{Z}^n$.

Theorem 2.3. *The parallelepiped* $P = \{u + t_1v_1 + t_2v_2 + \cdots + t_nv_n\colon u, v_1, v_2, \ldots, v_n \in \mathbf{Z}^n, 0 \le t_1, t_2, \ldots, t_n \le 1\}$ *is primitive if and only if volume*$(P) = 1$

3. PRIMITIVE TRIANGLES vs. PRIMITIVE TETRAHEDRA. We now compare primitive triangles in $\mathbf{Z}^2$ with primitive tetrahedra in $\mathbf{Z}^3$. Any primitive triangle in $\mathbf{Z}^2$ has area $\frac{1}{2}$. This is the simplest case of Pick's theorem.

Theorem 3.1 [Pick [7]]. *Let P be a lattice polygon in $\mathbf{Z}^2$. If there are I lattice points in the interior of P and B lattice points on the boundary of P, then area*$(P) = I + \frac{1}{2}B - 1$.

We refer the reader to **[1]** for two elementary and elegant proofs of Theorem 3.1; **[2]** and **[3]** are two recent expositions on Pick's theorem in the MONTHLY.

Reeve **[8]** observed that the tetrahedron with vertices $(0,0,0)$, $(1,0,0)$, $(0,1,0)$, and $(1,1,c)$ is primitive for any non-zero integer c. Consequently, there is no bound on the volume of a primitive tetrahedron and a direct generalization of Pick's theorem to $\mathbf{R}^3$ is impossible. However, by replacing $\mathbf{Z}^3$ with the *fractional* lattice $\frac{1}{n}\mathbf{Z}^3$ for a fixed positive integer n, he was able to find a three-dimensional analogue of Pick's theorem; we do not pursue this direction. In addition to **[8]** and **[9]**, the interested reader should look at **[6]**, where Reeve's theorem is extended to higher dimensions. Other related references are listed on the first page of **[2]**.

Another way to show that primitive triangles in $\mathbf{Z}^2$ have area equal to $\frac{1}{2}$ is to use Theorem 2.3. It is instructive to see how the argument proceeds and why it cannot be extended to give an upper bound on the volume of primitive tetrahedra.

Suppose $T = \{u + t_1v_1 + t_2v_2\colon u, v_1, v_2 \in \mathbf{Z}^2, 0 \le t_1, t_2 \le 1, t_1 + t_2 \le 1\}$ is a lattice point triangle containing the lattice point w. Let x be the coset of $\mathbf{Z}^2/(\mathbf{Z}v_1 \oplus \mathbf{Z}v_2)$ that w represents. *There is at least one lattice point in the triangle* $\{u + t_1v_1 + t_2v_2\colon u, v_1, v_2 \in \mathbf{Z}^2, 0 \le t_1, t_2 \le 1, t_1 + t_2 \ge 1\}$ *that represents the inverse* $-x$. It follows that T is primitive if and only if the associated parallelogram $P = \{u + t_1v_1 + t_2v_2\colon u, v_1, v_2 \in \mathbf{Z}^2, 0 \le t_1, t_2 \le 1\}$ is primitive. Using Theorem 2.3 we conclude that T is primitive if and only if *area*$(T) = \frac{1}{2}$.

Now let T_1 be the tetrahedron

$$T_1 = \{u + t_1v_1 + t_2v_2 + t_3v_3\colon u, v_1, v_2, v_3 \in \mathbf{Z}^3, 0 \le t_1, t_2, t_3 \le 1, t_1 + t_2 + t_3 \le 1\}.$$

The preceding argument shows that T_1 is primitive if and only if the tetrahedron

$$T_2 = \{u + t_1v_1 + t_2v_2 + t_3v_3\colon u, v_1, v_2, v_3 \in \mathbf{Z}^3, \\ 0 \le t_1, t_2, t_3 \le 1, 2 \le t_1 + t_2 + t_3 \le 3\}$$

is primitive. But we cannot conclude that the parallelepiped

$$P = \{u + t_1v_1 + t_2v_2 + t_3v_3\colon u, v_1, v_2, v_3 \in \mathbf{Z}^3, 0 \le t_1, t_2, t_3 \le 1\}$$

is primitive since, in addition to T_1 and T_2, there are four other tetrahedra inside P that need not be primitive. Therefore, we cannot find a bound for the volume of P and use it to give a bound on the volume of T_1.

Recall that the volume of any lattice point tetrahedron is $c/6$, where c is a positive integer. Suppose

$$T_1 = \{u + t_1v_1 + t_2v_2 + t_3v_3 \colon u, v_1, v_2, v_3 \in \mathbf{Z}^3, 0 \le t_1, t_2, t_3 \le 1, t_1 + t_2 + t_3 \le 1\}$$

is a primitive tetrahedron and $vol(T_1) = c/6$ with $c \ge 2$. The parallelepiped

$$P = \{u + t_1v_1 + t_2v_2 + t_3v_3 \colon u, v_1, v_2, v_3 \in \mathbf{Z}^3, 0 \le t_1, t_2, t_3 \le 1\}$$

has volume c. By Theorem 2.2 there are $c - 1$ non-zero cosets in the quotient group $\mathbf{Z}^3/(\mathbf{Z}v_1 \oplus \mathbf{Z}v_2 \oplus \mathbf{Z}v_3)$. *It now follows that there are $c - 1$ lattice points in the interior of P and each one represents a distinct coset of* $\mathbf{Z}^3/(\mathbf{Z}v_1 \oplus \mathbf{Z}v_2 \oplus \mathbf{Z}v_3)$. This assertion is based on the following two observations:

1. A coset has a unique representative in P if and only if the representative lies in the interior of P.
2. There are no lattice points on the boundary of P other than the vertices. Suppose $w = u + t_1v_1 + t_2v_2 + t_3v_3$ is a lattice point lying on the boundary of P that is not a vertex of P. There are two possible cases. Case 1: Exactly 1 member of the set $\{t_1, t_2, t_3\}$, say t_3, is an integer. In this case, either the lattice point $u + t_1v_1 + t_2v_2$, or the lattice point $u + (1 - t_1)v_1 + (1 - t_2)v_2$ lies on one of the faces of T_1. Since $t_1, t_2 \notin \mathbf{Z}$, neither lattice point is a vertex of P. Case 2: Exactly 2 members of the set $\{t_1, t_2, t_3\}$, say t_2 and t_3, are integers. In this case, the lattice point $u + t_1v_1$ lies on one of edges of T_1. Since $t_1 \notin \mathbf{Z}$, $u + t_1v_1$ is not a vertex of T_1. Therefore, if P contains a lattice point on its boundary that is not a vertex, then T_1 is not primitive.

Thus there are $c - 1$ lattice points in the interior of P. How are these $c - 1$ lattice points arranged inside P? The answer to this question gives us a geometric characterization of primitive tetrahedra in $\mathbf{Z}^3$.

Theorem 3.2. *Consider the tetrahedron*

$$T = \{u + t_1v_1 + t_2v_2 + t_3v_3 \colon u, v_1, v_2, v_3 \in \mathbf{Z}^3, 0 \le t_1, t_2, t_3 \le 1, t_1 + t_2 + t_3 \le 1\}$$

and the parallelepiped

$$P = \{u + t_1v_1 + t_2v_2 + t_3v_3 \colon u, v_1, v_2, v_3 \in \mathbf{Z}^3, 0 \le t_1, t_2, t_3 \le 1\}.$$

Then T is primitive if and only if all the non-vertex lattice points in P lie in the interior of one of the three diagonal parallelograms of P that do not intersect the interior of T, i.e., one of the parallelograms with vertices $\{u + v_1, u + v_2, u + v_1 + v_3, u + v_2 + v_3\}$, $\{u + v_2, u + v_3, u + v_2 + v_1, u + v_3 + v_1\}$, *or* $\{u + v_3, u + v_1, u + v_3 + v_2, u + v_1 + v_2\}$.

The sufficiency of the stated criterion is trivial and all the work lies in proving its necessity. We have been unable to construct a direct geometric proof and we pose this as a problem to our readers. We prove Theorem 3.2 by proving an analytic version, Theorem 4.2.

4. UNIMODULAR MAPS.

Before we can describe an analytic version of Theorem 3.2, we need to define the concept of a unimodular map.

Definition 4.1. The map $f \colon \mathbf{R}^n \to \mathbf{R}^n$ is a *unimodular map* if (i) f is affine, (ii) f preserves volume, and (iii) f maps points in $\mathbf{Z}^n$ to points in $\mathbf{Z}^n$.

We leave it to the reader to check the following properties of unimodular maps.

Theorem 4.1. *Let $f: \mathbf{R}^n \to \mathbf{R}^n$ be a map and let P be a lattice polytope in $\mathbf{Z}^n$.*

(i) *f is unimodular if and only if $f(u) = Mu + v$, where $M \in GL_n(\mathbf{Z})$, i.e., M is a $n \times n$ matrix with integer entries and* $\det(M) = \pm 1$, *and $v \in \mathbf{Z}^n$.*
(ii) *If f is unimodular, then f is invertible and f^{-1} is unimodular.*
(iii) *If f is unimodular then $f(P)$ and $f^{-1}(P)$ are lattice polytopes. Furthermore, f maps interior points of P to interior points of $f(P)$, boundary points of P to boundary points of $f(P)$, and vertices of P to vertices of $f(P)$.*
(iv) *If P is primitive, then both $f(P)$ and $f^{-1}(P)$ are primitive.*

Unimodular maps define an equivalence relation on the set of lattice polytopes and so we can consider equivalence classes of lattice polytopes.

Definition 4.2. Two lattice polytopes P_1, P_2 in $\mathbf{R}^n$ are *unimodularly equivalent* if there is a unimodular map $f:\mathbf{R}^n \to \mathbf{R}^n$ such that f maps P_1 to P_2. If P_1, P_2 are unimodularly equivalent we write $P_1 \cong P_2$.

Any two primitive triangles in $\mathbf{Z}^2$ are unimodularly equivalent. One way to prove this is to apply Theorem 2.2. Suppose the triangles

$$T_1 = \{u_1 + t_1 v_1 + t_2 v_2 : u_1, v_1, v_2 \in \mathbf{Z}^2, 0 \le t_1, t_2 \le 1, t_1 + t_2 \le 1\}$$

and

$$T_2 = \{u_2 + t_1 w_1 + t_2 w_2 : u_2, w_1, w_2 \in \mathbf{Z}^2, 0 \le t_1, t_2 \le 1, t_1 + t_2 \le 1\}$$

are primitive. Then the associated parallelograms

$$P_1 = \{u_1 + t_1 v_1 + t_2 v_2 : u_1, v_1, v_2 \in \mathbf{Z}^2, 0 \le t_1, t_2 \le 1\}$$

and

$$P_2 = \{u_2 + t_1 w_1 + t_2 w_2 : u_2, w_1, w_2 \in \mathbf{Z}^2, 0 \le t_1, t_2 \le 1\}$$

are primitive. Let A be the 2×2 matrix whose first column is v_1 and whose second column is v_2; and let B be the 2×2 matrix whose first column is w_1 and whose second column is w_2. Since P_1 and P_2 are primitive parallelograms, $A, B \in GL_2(\mathbf{Z})$. The unimodular map $f(x) = BA^{-1}x + (u_2 - BA^{-1}u_1)$ maps T_1 to T_2 (and P_1 to P_2). This argument can be used to show that any two primitive parallelepipeds in $\mathbf{Z}^n$ are unimodularly equivalent.

We now describe an analytic version of Theorem 3.2, which appears as Corollary 5.7 in **[10]**. From here on, $T_{a,b,c}$ denotes the tetrahedron with vertices $(0,0,0)$, $(1,0,0)$, $(0,1,0)$, and (a,b,c), with $(a,b,c) \in \mathbf{Z}^3$ and $c \neq 0$.

Theorem 4.2 [Reeve-White-Howe-Scarf-Reznick]. *A tetrahedron T is primitive if and only if $T \cong T_{0,0,1}$ or $T \cong T_{1,b,c}$ with $1 \le b < c$ and* $\gcd(b,c) = 1$.

We refer the reader to **[10]** for a proof of Theorem 4.2, and to **[11]** and **[13]** for proofs of variants of it. It is surprising that such an elegant and simple theorem is not better known.

5. THE COUNTING FORMULA. Unimodular maps preserve volume; consequently, the number of distinct (unimodular) equivalence classes of primitive tetrahedra is infinite. We are led to consider the following question.

For a given positive integer c, is there a counting formula for the number of distinct (unimodular) equivalence classes of primitive tetrahedra of volume $c/6$? The answer lies in determining the relationship between the integers x and b when $T_{1,b,c} \cong T_{1,x,c}$. The relationship is described in the following variant of Theorem 5.6 in **[10]**.

Theorem 5.1. *Let $b, b^{-1}, c, x \in \mathbf{Z}$ with $1 \le x, b, b^{-1} < c$, $\gcd(b,c) = \gcd(x,c) = 1$, and $bb^{-1} \equiv 1 (\operatorname{mod} c)$. The primitive tetrahedra $T_{1,b,c}$ and $T_{1,x,c}$ are unimodularly equivalent if and only if $x \in \{b, c-b, b^{-1}, c-b^{-1}\}$.*

Proof: ($\Rightarrow$) We can view a unimodular map from $T_{1,b,c}$ to $T_{1,x,c}$ as a map between the vertices of $T_{1,b,c}$ and $T_{1,x,c}$. There are 24 possible maps. An examination of each of these shows that $x \in \{b, c-b, b^{-1}, c-b^{-1}\}$.

For example, suppose we look at the unimodular map

$$(0,0,0) \mapsto (0,1,0), (1,0,0) \mapsto (1,x,c), (0,1,0) \mapsto (0,0,0), (1,b,c) \mapsto (1,0,0).$$

This map has the form

$$\begin{pmatrix} 1 & 0 & 0 \\ x-1 & -1 & a \\ c & 0 & -1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

with the conditions that $a \in \mathbf{Z}$ and $x - b + ac = 0$. Since $|x - b| < c$, it follows that $a = 0$ and therefore $x = b$.

($\Leftarrow$) $T_{1,b,c} \cong T_{1,c-b,c}$ via the unimodular map

$$\begin{pmatrix} 1 & 0 & 0 \\ -1 & -1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},$$

$T_{1,b,c} \cong T_{1,b^{-1},c}$ via the unimodular map

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & b^{-1} & \dfrac{1 - bb^{-1}}{c} \\ 0 & c & -b \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},$$

and $T_{1,b,c} \cong T_{1,c-b^{-1},c}$ via the unimodular map

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & c - b^{-1} & -\dfrac{b(c - b^{-1}) + 1}{c} \\ 0 & c & -b \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}. \qquad \blacksquare$$

Before proceeding to the statement of the counting formula, it is instructive to determine the number of equivalence classes of primitive tetrahedra of volume $c/6$ for some specific values of c. We do so for $c = 13$ and $c = 14$. We start with the case $c = 13$. Because of Theorem 4.2 we need consider only the 12 primitive tetrahedra $T_{1,1,13}, T_{1,2,13}, T_{1,3,13}, \ldots, T_{1,12,13}$ and determine which are unimodularly equivalent. Theorem 5.1 ensures that $T_{1,1,13}, \cong T_{1,12,13}$, $T_{1,2,13} \cong T_{1,6,13} \cong T_{1,7,13} \cong T_{1,11,13}$, $T_{1,3,13} \cong T_{1,4,13} \cong T_{1,9,13} \cong T_{1,10,13}$, and $T_{1,5,13} \cong T_{1,8,13}$, so there are 4 equivalence classes of primitive tetrahedra of volume 13/6. For $c = 14$, we need to consider only the 6 primitive tetrahedra $T_{1,1,14}$, $T_{1,3,14}$, $T_{1,5,14}$, $T_{1,9,14}$,

$T_{1,11,14}$, and $T_{1,13,14}$. Since $T_{1,1,14} \cong T_{1,13,14}$ and $T_{1,3,14} \cong T_{1,5,14} \cong T_{1,9,14} \cong T_{1,11,14}$, there are 2 equivalent classes of primitive tetrahedra of volume 14/6.

We now prove the counting formula by combining Theorem 5.1 with Burnside's lemma.

Theorem 5.2 (Burnside's lemma). *Let a finite group G act on a finite set X and have N orbits. Then*

$$N = \frac{1}{|G|} \sum_{g \in G} |X_g|,$$

where $X_g = \{x\colon g(x) = x,\ x \in X\}$, i.e., X_g is the set of fixed points of g.

See [**12**, Theorem 2.71, p. 125] for a proof, as well as an explanation of why it is referred to by some as *not-Burnside's* lemma.

We are now ready to state and prove the counting formula.

Theorem 5.3. *Suppose n is an integer greater than 2, $T(n)$ is the number of distinct equivalence classes of primitive tetrahedra of volume $n/6$, $f(x) \in (\mathbf{Z}/n\mathbf{Z})[x]$, and $N(f(x) \equiv 0 \pmod n)$ denotes the number of solutions of the congruence $f(x) \equiv 0 \pmod n$ in $\mathbf{Z}/n\mathbf{Z}$. Then*

$$T(n) = \frac{\phi(n) + N(x^2 - 1 \equiv 0 \pmod n) + N(x^2 + 1 \equiv 0 \pmod n)}{4}.$$

In particular, if n is a prime p, then

$$T(p) = \left[\frac{p}{4}\right] + 1 = \begin{cases} \dfrac{p+3}{4} & \text{if } p \equiv 1 \pmod 4 \\[2mm] \dfrac{p+1}{4} & \text{if } p \equiv 3 \pmod 4. \end{cases}$$

Proof: Let $U = \{x\colon 1 \le x < n,\ \gcd(x, n) = 1\}$ be the set of units of $\mathbf{Z}/n\mathbf{Z}$ and let $G = \{g_1, g_2, g_3, g_4\}$ be the group of bijections of U, where $g_1(x) = x$, $g_2(x) = n - x$, $g_3(x) = x^{-1}$, and $g_4(x) = n - x^{-1}$ for any element $x \in U$. The cardinality of U is $\phi(n)$, the Euler phi function evaluated at n, and $G \cong \mathbf{Z}/2\mathbf{Z} \oplus \mathbf{Z}/2\mathbf{Z}$.

Theorem 5.2 says that $T(n)$ equals the number of orbits of U under the action of G. Let $U_i = \{x\colon g_i(x) = x,\ x \in U\}$, $i = 1, 2, 3, 4$. Since g_1 is the identity it fixes all of U. When $n > 2$, g_2 has no fixed points. The fixed points of g_3 are the solutions of the congruence $x^2 \equiv 1 \pmod n$ and the fixed points of g_4 are the solutions of the congruence $x^2 \equiv -1 \pmod n$. Thus,

$$|U_1| = \phi(n),\ |U_2| = 0,\ |U_3| = N(x^2 - 1 \equiv 0 \pmod n),$$
$$|U_4| = N(x^2 + 1 \equiv 0 \pmod n).$$

We now apply Burnside's lemma to obtain the formula

$$T(n) = \frac{\phi(n) + N(x^2 - 1 \equiv 0 \pmod n) + N(x^2 + 1 \equiv 0 \pmod n)}{4}. \qquad \blacksquare$$

Theorem 4.2 shows that if T is a primitive tetrahedron of volume 1/6 then $T \cong T_{0,0,1}$, and if T is a primitive tetrahedron of volume 1/3 then $T \cong T_{1,1,2}$; so in the exceptional cases when $n = 1, 2$ we have $T(n) = 1$. The reader can observe

that $\phi(n)$, $N(x^2 - 1 \equiv 0 \pmod n)$, and $N(x^2 + 1 \equiv 0 \pmod n)$ are multiplicative functions of n. We now cite a result from the theory of quadratic residues that gives a formula for $N(x^2 - 1 \equiv 0 \pmod n)$ and $N(x^2 + 1 \equiv 0 \pmod n)$.

Theorem 5.4. *Let $n = 2^k m$, with $n \geq 2$ and $\gcd(m, 2) = 1$. If $m > 1$, let the prime divisors of m be $p_1, \ldots, p_t$; otherwise set $t = 0$. Then*

$$N\big(x^2 - 1 \equiv 0 \pmod n\big) = \begin{cases} 2^t & \text{if } k = 0, 1 \\ 2^{t+1} & \text{if } k = 2 \\ 2^{t+2} & \text{if } k \geq 3, \end{cases}$$

and

$$N\big(x^2 + 1 \equiv 0 \pmod n\big)$$
$$= \begin{cases} 2^t \text{ if } k \leq 1 \text{ and } p_i \equiv 1 \pmod 4 \text{ for any } i \in \{1, \ldots, t\} \\ 0 \text{ if } k \geq 2 \text{ or there exists an } i \in \{1, \ldots, t\} \text{ such that } p_i \equiv 3 \pmod 4. \end{cases}$$

See [**5**, Chapter 2, Section 8; Chapter 3, Section 5] for a proof.

6. HOWE'S THEOREM. We now describe Howe's theorem on primitive polyhedra. Theorem 2.1 shows that any primitive polyhedron in $\mathbf{R}^3$ has at most 8 vertices. Howe proved that any primitive polyhedron with eight vertices is, up to a unimodular map, the convex hull of a square and a parallelogram.

Theorem 6.1 [Howe]. *Let P be a primitive polyhedron with eight vertices. Then there is a unimodular map that maps P to the polyhedron whose vertices are $(0,0,0)$, $(1,0,0)$, $(0,1,0)$, $(0,0,1)$, $(0,1,1)$, $(1,a,b)$, $(1,c,d)$, and $(1, a+c, b+d)$ with $a, b, c, d, \in \mathbf{Z}$, $a, b, c, d \geq 0$ and $ad - bc = 1$. Furthermore, any primitive polyhedron with fewer than eight vertices can be embedded in one with eight vertices.*

We refer the reader to the article by Scarf [**13**] for a proof of this theorem. It should be noted that the four vertices $(0,0,0)$, $(0,1,0)$, $(0,0,1)$, and $(0,1,1)$ lie on the plane $x = 0$ and form a square of area 1. The other four vertices $(1,0,0)$, $(1,a,b)$, $(1,c,d)$, and $(1,a+c,b+d)$ lie on the parallel plane $x = 1$ and form a parallelogram of area 1.

The problem of characterizing primitive tetrahedra was independently studied in 1957 by Reeve [**8**] and in 1964 by White [**15**]; Theorem 4.2 arises from combining their results. In 1977, Howe independently discovered Theorem 4.2 and its generalization, Theorem 6.1. He did not publish his work and it was Scarf [**13**] who first publicized Howe's theorem. Over the years, other mathematicians have rediscovered Theorem 4.2. For example, Therese Hart, Karen Rogers, and I discovered it in 1991 and wrote up our results in the unpublished manuscript [**4**]. We then came across Reznick's [**10**] article where we learnt of the work of Reeve, White, and others. The contents of [**4**] are included in Chapter 2 of K. Rogers' doctoral dissertation [**11**]. The last chapter of this dissertation contains some partial results on primitive simplices in $\mathbf{Z}^4$.

We leave the reader with the following question: Are there analogues of the counting formula for primitive polyhedra with 5, 6, 7, and 8 vertices?

ACKNOWLEDGMENTS. I am grateful to K. Rogers for showing me her proof of Theorem 4.2 and some of the unimodular equivalences of $T_{a,b,c}$. The use of Burnside's lemma to prove the counting formula was suggested by a referee and is an undoubted improvement on my initial proof. The referees and my friend Sergio Alvarez made extensive suggestions on improving the exposition.

REFERENCES

1. D. DeTemple, Pick's formula: a retrospective, *Mathematics Notes from Washington State University*, Vol. 32, Nos.3-4 (November, 1989).
2. R. Diaz and S. Robins, Pick's formula via the Weierstrass $\wp$-function, *Amer. Math. Monthly* **102** (1995) 431–437.
3. B. Grunbaum and G. C. Shephard, Pick's theorem, *Amer. Math. Monthly* **100** (1993) 150–161.
4. T. Hart, M. Khan, and K. Rogers, Primitive tetrahedra in $\mathbf{Z}^3$ (unpublished manuscript).
5. Hua Loo Keng, *Introduction to Number Theory*, translated from the Chinese by P. Shiu, Springer-Verlag, 1982.
6. I. G. MacDonald, The volume of a lattice polyhedron, *Proc. Camb. Phil. Soc.* **59** (1963) 719–726.
7. G. Pick, Geometrisches zur Zahlenlehre, *Sitzungber. Lotos* (Prague) **19** (1899) 311–319.
8. J. E. Reeve, On the volume of lattice polyhedra, *Proc. London Math. Soc.* (3) **7** (1957) 378–395.
9. J. E. Reeve, A further note on the volume of lattice polyhedra, *J. London Math. Soc.* **34** (1959) 57–62.
10. B. Reznick, Lattice point simplices, *Discrete Mathematics* **60** (1986) 219–242.
11. K. M. Rogers, Primitive simplices in $\mathbf{Z}^3$ and $\mathbf{Z}^4$ (Ph.D. dissertation), Columbia University, 1993.
12. J. Rotman, *A first course in abstract algebra*, Prentice-Hall, 1996.
13. H. E. Scarf, Integral polyhedra in three space, *Math. Oper. Res.* **10** (1985) 403–438.
14. C. L. Siegel, *Lectures on the Geometry of Numbers*; notes by B. Friedman; rewritten by K. Chandrasekharan with the assistance of R. Suter, Springer-Verlag, 1989.
15. G. K. White, Lattice tetrahedra, *Canad. J. Math.* **16** (1964) 389–396.

MIZAN R. KHAN received his B.Sc. and M.Sc. in Mathematics from the London School of Economics under the tutelage of Haya Freedman. He received his Ph.D. from the University of Massachusetts, Amherst, where he wrote a dissertation in computational number theory under the direction of David Hayes. Before coming to Eastern Connecticut State University, he held visiting positions at Mt. Holyoke College and Trinity College.
Department of Mathematics and Computer Science, Eastern Connecticut State University, Willimantic, CT 06226
khanm@ecsuc.ctstateu.edu

What Makes a Great Mathematics Teacher? The Case of Augustus De Morgan

Adrian Rice

1. INTRODUCTION. It is often said that no one forgets a good teacher. Whether this statement is true or not, almost everyone can recall at least one teacher who influenced some aspect of his or her future study or career. But there are occasional examples of *great* mathematics teachers who instill a remarkable number of their students with a love and enthusiasm for the subject, which has a lasting and profound effect on them, even if they never become practising mathematicians.

The nineteenth century British mathematician Augustus De Morgan (1806–1871) was one such teacher. Although his name is well known to any student of set theory, his chief mathematical legacy arose from his novel research in logic. This research created the first logic of relations, and promoted a symbolic approach to the subject, in which regard he greatly encouraged the work of his friend and contemporary George Boole. [**34**] De Morgan was also interested in algebra [**26**], and his attempts to extend the geometrical representation of complex numbers influenced the discovery of quaternions by his friend William Rowan Hamilton. Mathematical analysis, a subject very much neglected in early nineteenth-century Britain, also occupied much of De Morgan's attention, and he produced notable work on convergence of series. [**20**, pp. 148–9] He also published many research papers on various aspects of the history of mathematics, about which he was a rare authority at the time. [**29**], [**31**]

By the end of his career in the mid-1860s, De Morgan was one of the most influential and highly-regarded mathematicians in Britain, outliving Hamilton and Boole by several years and out-ranking the likes of Cayley and Sylvester (themselves far more original mathematicians) by virtue of his age. But how was this reputation achieved? Why was he so highly regarded?

One obvious reason is curiously often the most overlooked. For virtually his entire career, De Morgan was professor of mathematics at University College London, a radically innovative establishment, which, at its foundation in 1826, was the first university-level institution to be established in England since Oxford and Cambridge in the Middle Ages. There, he single-handedly delivered courses on mathematics to a generation of undergraduates for a third of a century. Because he was in charge of mathematical tuition at the leading higher educational institutional in his nation's capital, he was a formative influence on numerous mathematicians, scientists, and other prominent intellectual figures of the Victorian period.

Unfortunately, there is a pronounced absence of published material pertaining to De Morgan's work as a teacher. We know that he wrote a series of very successful textbooks, and these are useful to some extent; but they shed little light on what he actually taught in his lecture room.

However, there is a major unpublished source of information about De Morgan's teaching. This source was described in the *Encylopœdia Britannica* as "a large mass

of mathematical tracts which he prepared for the use of his students, treating all parts of mathematical science, and embodying some of the matter of his lectures". [**24**, p. 9] They are preserved in the University of London Library in the form of more than 320 notebooks containing the majority of De Morgan's course material in his own handwriting. The contents of these manuscripts give us a considerable insight into the material covered by mathematics students at the most progressive educational establishment in nineteenth-century England.

In this article, we first investigate the content of De Morgan's mathematical course and examine his teaching methodology; we then consider evidence from his students about how this material came across. Finally, by comparing his teaching style with that of other well known nineteenth-century mathematicians, we highlight some of the ingredients that go into making a great mathematics teacher, and show that these criteria are satisfied in the case of Augustus De Morgan.

2. DE MORGAN'S PROFESSORIAL CAREER. Following his birth in India in June 1806, De Morgan spent the majority of his formative years in southwest England, where he received an adequate classical education. In 1823, aged only sixteen, he entered Trinity College, Cambridge, where his mathematical talents were nurtured by his tutors, who included the prominent English mathematician George Peacock. Peacock, Charles Babbage, and John Herschel had founded the short-lived but influential Analytical Society in 1812, [**17**] which helped secure the adoption of Lagrange's algebraic methods of calculus in the Cambridge syllabus, replacing the Newtonian fluxional system, which had been entrenched in Britain for well over a century. [**21**]

De Morgan's Cambridge years coincided with the foundation of a university in London, the only capital city in Europe without such an institution at this time. Indeed, up to this period, Oxford and Cambridge were the only places in England to offer university qualifications, and since they were fully open only to members of the Church of England, denominations such as Catholics or Jews were effectively barred from university degrees. So too were the urban middle classes who, while not poor, were nevertheless financially incapable of supporting their offspring through courses of study away from home. The establishment of University College London (originally titled "London University") in 1826 was a radical solution to this problem, made all the more so by its explicit secular character and progressive programme of studies.

Equally daring was its choice of founding professor of mathematics. De Morgan was appointed to the position in February 1828, scarcely a year after his graduation from Cambridge, aged only twenty-one. [**30**] But not all went smoothly. After opening for lectures the following October, the new university was plagued by financial troubles and petty personal disputes. After a professorial colleague was dismissed in 1831, De Morgan immediately resigned on principle. Five years later, however, he was invited to return after the premature death of his replacement. He was to remain for a further thirty years.

His final departure was occasioned by the college's (non-)adherence to its policy of religious equality; indeed, he remains the only professor in the history of University College to have resigned *twice* on matters of principle. For De Morgan, the college's refusal to appoint a candidate to the vacant chair of philosophy on the grounds of his being a controversial Unitarian minister was a betrayal of its founding principles. He resigned his professorship on 10 November 1866, giving his last lecture in the summer of 1867. He never returned, refusing even a request from his former students to sit for a bust to be placed in the college library,

Figure 1. Augustus De Morgan pictured in 1866.

explaining that, as far as he was concerned, "our old college no longer exists". [**15**, p. 360]

3. THE COURSE. During the period of De Morgan's professorship, his mathematics course formed a central component of the college's curriculum. It was intended to constitute part of the students' first two years, during which time they would also study such subjects as Latin, Greek, and natural philosophy (i.e., physics). Since school education was not yet compulsory and the school leaving age was, on average, around fourteen, the students at University College in De Morgan's day were substantially younger than they are now. In general, they ranged from 15 to 18, usually leaving the college to begin vocational training, employment or, in the case of the exceptional students, more advanced study at Oxford (if they were classically inclined) or Cambridge (if mathematically).

In mathematics the students were divided into classes corresponding to the first and second years of undergraduate study, with each class being further divided into lower and higher divisions. The first year (or junior) course was designed to contain "what is most essential for those who are intended for practical professions, such as Civil Engineers, &c.", [**33**, p. 42] while the senior class was intended for those capable of tackling more advanced topics. However, due to the range of materials available, the course was "confined principally to those parts of the subject which are necessary for the study of Natural Philosophy". [**33**, p. 42] The following outline indicates that this definition was a very broad one:

JUNIOR CLASS, LOWER DIVISION:

i) Arithmetic and the arithmetical theory of proportion
ii) Euclid, Books 1–4
iii) 6th Book of Euclid
iv) First book of Solid Geometry in Euclid
v) Algebra, arithmetically considered, up to equations of the first degree.

JUNIOR CLASS, HIGHER DIVISION:

i) Euclid, Books 5 and 6
ii) First book of Solid Geometry in Euclid
iii) A review of the principles and operations of arithmetic
iv) Algebra (including the nature and use of logarithms)
v) Plane trigonometry (including mensuration).

SENIOR CLASS, LOWER DIVISION:

i) Spherical trigonometry
ii) Conic sections
iii) Applications of algebra to geometry
iv) Higher parts of algebra
v) Differential and integral calculus.

SENIOR CLASS, HIGHER DIVISION:

> Extension of subjects in the Lower Senior Class. "Subjects which all must learn who wish to become analysts, whether for Engineering or any other pursuit." [**39**, pp. 6–7], [**40**, pp. 7–8], [**41**, pp. 19, 35]

However, De Morgan was at pains to point out that this plan should not be regarded as a definitive declaration of intentions. As he said in his inaugural lecture of 1828, "I shall not consider myself bound to carry the class through the whole of what is contained in it if it shall appear that their interest will be more effectually consulted by my confining myself to the more prominent parts of it." [**6**, f. 45] As far as he was concerned, it was quality of knowledge that mattered more than quantity.

In order to achieve this, De Morgan highlighted the two principal methods whereby his students could acquire mathematical knowledge: "The first is by diligent study in the retirement of the closet; the second, by haunting the benches of the lecture-room, and picking up what may chance to fall." [**13**, p. 14] Lectures alone, he maintained, were insufficient to bring the student to the appropriate level of understanding. Moreover, student lecture notes, while important, were no substitute for a full treatise; indeed, he compared the information obtained from listening to a lecture to the comprehension achieved from reading a book at speed.

De Morgan thus regarded the role of lectures as merely providing students with assistance in difficulty and guidance on relevant reading.

In order to enlarge this oral instruction, he prepared a vast quantity of handwritten tracts on all aspects of his course, which were then placed in the University College library for his students to refer to. They were designed to supplement not only the lecture material, but also the wide reading that De Morgan expected his students to undertake.[1] The surviving notebooks (written between 1843 and 1866) are well over three hundred in number, each featuring De Morgan's legible handwriting. De Morgan also had an idiosyncratic habit of pasting in printed material that he considered particularly relevant; these insertions often consisted of an appropriate paper, usually by himself. These tracts reveal that he taught much more than was indicated in the published syllabi and exam papers. What now follows is a survey of the material contained in the existing tracts to give some idea of what a student of mathematics at University College could have expected to study under De Morgan 150 years ago.

TABLE 1. De Morgan's surviving tracts

CLASS	NUMBER OF NOTEBOOKS
Lower Junior	10
Higher Junior	70
Lower Senior	109
Higher Senior	138

3.1 The Lower Junior Class. Of the 327 surviving notebooks, only ten contain material designed for the use of students in De Morgan's lower junior class. Not only was the subject matter far less extensive than in succeeding classes, but also, in the mode of tuition adopted by the Professor for this class, oral lectures occupied a very small place, the majority of the time being devoted to giving written exercises and answering students' questions. Furthermore, for much of the relevant material at this introductory level, existing textbooks were perfectly adequate, such as his own *Elements of Arithmetic* (1830) and *Elements of Algebra* (1835), as well as numerous editions of Euclid's *Elements*.

As well as giving alternative presentations of material that could be found in the students' books, the tracts dealt in some considerable depth with matters with which most textbooks (even De Morgan's) did not concern themselves. One of the most fascinating tracts (#110–comprising three notebooks) was designed to be read before the student opened the first page of Euclid. Entitled "Notions preliminary to Geometry", it illustrates De Morgan's desire for his students to be acquainted from the very start with the philosophical and epistemological issues relating to the subject. More significantly, it demonstrates his belief that a thorough grounding in logical notions and processes was essential for the students'

[1] In the absence of photocopiers, many would copy the contents of these tracts wholesale, as evinced by a large volume in University College Library containing a student's transcription of 33 of them: University College London Archives, MS.ADD.6, "Mathematical Tracts by Professor De Morgan, copied from the original Manuscripts in the Library of University College London by John Power Hicks 1849–1851".

understanding of geometrical arguments.[2] As he said:

> The principles on which geometrical propositions are established belong to the totally distinct and equally simple science of logic; and since geometry without logic would be absurd, it is desirable that the principles of the latter science should be studied with precision previously to employing them upon the former. [**9**, pp. 238–9]

De Morgan was wont to complain about the lack of contact between the disciplines of mathematics and logic: "Geometers have seldom been very *formal* logicians; and their patent of exemption was signed by Euclid." [**7**, p. 435] One of the principal sources of confusion when initiating students into the study of geometrical demonstrations was the distinction between a proposition and its converse. So, for example, the statement that 'all equilateral triangles are equiangular' was often taken to imply that 'all equiangular triangles are equilateral'. "These errors," said De Morgan, "should be guarded against beforehand, by exercising the pupil in simple deductions, such as are to be found in every syllogism, taking care that all terms used have reference to objects with which they are familiar. It should be illustrated to them that the truth of an argument depends on two distinct considerations, the truth of the premises, and the manner in which the conclusion is deduced from them." [**8**, pp. 272–3]

The next step before coming to actual geometrical demonstrations was to introduce his students to the concept of a proof. "A proposition", he wrote, "may be proved in two ways: Directly, by showing that it *is true*. Indirectly, by showing that the contradiction *is false*." [**46**, f. 8] Since the latter was conceptually the most difficult for the beginner, this tract was principally concerned with this mode of procedure, which "forces an absurd result out of the contradiction, and therefore forces the denial of the contradiction, or the affirmation of the proposition". [**46**, f. 9] Proof by contradiction, was, in De Morgan's opinion, rendered far more intelligible by the early study of logic, and mastery of the technique was a vital skill to acquire before tackling Euclid.

3.2 The Higher Junior Class. On entering De Morgan's higher junior class, his students were expected to be fully familiar with Euclidean deductive reasoning up to the fourth book of the *Elements*. But the fifth book, introducing the complex ideas of ratios and proportion, often caused the most problems. De Morgan noted that, owing to its highly convoluted presentation, "it has been customary for mathematical students among us to read the Fifth Book of Euclid; frequently without understanding it". [**10**, p. iii] For this reason, he substituted arithmetical notions of proportion instead of the traditional geometrical ones, a simplification that helped to make the subject far more intelligible to his junior students than if he had left them to study it unaided.

Despite his almost instinctive mathematical abilities, De Morgan was fully aware of the need to eliminate as many barriers as possible to the beginner's

[2] His writings on geometrical education provide the first published evidence of De Morgan's interest in logic, although at this point it was utilised purely as a pedagogic tool. He later elaborated his ideas in a short book for his students entitled *First Notions of Logic(preparatory to the study of geometry)*, published in 1839. This was later incorporated as the first chapter of his *Formal Logic* in 1847, by which time his interest in logic had transcended its utility merely as an aid to geometry, and was manifesting itself in the publication of research papers concentrating more on the intrinsic nature of the subject itself.

understanding of unfamiliar mathematical topics. A further aid to the students' geometric cognition was his rejection of perspective drawings in favour of three-dimensional models. As he explained to the audience of his introductory lecture:

> Considerable obstacles generally present themselves to the beginner, in studying the elements of Solid Geometry, from the practice which has hitherto uniformly prevailed in this country, of never submitting to the eye of the student, the figures on whose properties he is reasoning, but of drawing perspective representations of them upon a plane. ... I hope that I shall never be obliged to have recourse to a perspective drawing of any figure whose parts are not in the same plane. [**6**. ff. 50–1]

Having dwelt extensively on Euclidean-related matters, the higher junior class would have then turned their attention to a recapitulation of the rules and procedures of arithmetic before being initiated into abstract algebra. The lower junior class would have already practised linear equations, but these were treated more as features of a universal arithmetic than a general algebra. Now the class was ready to learn the distinction between the two: "In *Arithmetic* every symbol of magnitude ... represents a *number*, and nothing but a number ... *Algebra* employs the *symbols*, the *language* and the *rules* of arithmetic. But ... no letter *a* has its full meaning described until we are told both its *value* and its *sign*." [**50**, ff. 1, 4]

De Morgan's algebra was a long way from the highly-evolved structural algebra of today. Yet it was to play a significant role in the development of abstract algebra, building on earlier work by his friend and former tutor, George Peacock. [**27**] However, whereas in Peacock's algebra the symbols were generally understood to represent numbers or operations, De Morgan would deliberately keep them abstract. "Thus," he wrote "*addition* is to be, for the present, a sound void of sense. It is a mode of combination represented by +; when + receives its meaning, so also will the word *addition*." [**14**, p. 101] It was an area to which he was to devote much research, although, since it was too advanced for his junior classes, he deferred its discussion until his pupils had reached the senior level.

Now that they were familiar with algebraic terminology and ideas, the students were ready to progress to the solution of quadratic, cubic, and higher order equations. It was at this stage that students would have first come across the binomial theorem, leading immediately to the study of series, both finite and infinite. This turn led them to convergent and divergent series, resulting in their introduction to one of the most crucial mathematical concepts, recently reinstated in analysis: the limit.

In Cambridge during the second decade of the nineteenth century, the Analytical Society had been instrumental in replacing Newton's fluxional calculus with the algebraic method of Lagrange, [**16**] thus rejecting a system based (albeit very dubiously) on the notion of a limit. This concept had been reformulated by Cauchy in the early 1820s, but was not immediately accepted in France or elsewhere. De Morgan's *Elements of Algebra* was the first English work to contain a definition of the continuity of a mathematical function using limits. His subsequent treatise on *The Differential and Integral Calculus* (1842), the most comprehensive English work on the subject for over a generation, was entirely grounded on the concept of limits.

But, he said, it was meaningless to ask what a limit was since "'What *is* the limit' is the same question as 'What is the exact expression for that which does not admit of exact expression'." [**51**, f. 2] What then was the point of introducing limits

in the first place? "We introduce them because we cannot do without them, being all the time perfectly willing to do without them if any one will show us how." [**51**, f. 3] Yet despite these reservations, De Morgan was nonetheless firmly convinced of their epistemological soundness and, as one of the first in Britain to publish and teach mathematics using limits, his work helped establish this concept as the basis of modern mathematical analysis.

The use of limits also formed the basis of his introduction to logarithms as well as his teaching of trigonometrical analysis, where he employed limits to analyse the various properties of expressions such as sin x/x and $(1 - \cos x)/x^2$. His tracts on trigonometry for this class began with the usual problems of plane trigonometry such as finding values of angles and sides given certain information, progressing to questions involving multiple angles and inverse functions.

These were the topics studied by De Morgan's higher junior class, as specified by the published syllabus. However, the existing tracts reveal that students were also given instruction in other areas, the first being interest and annuities. Interest, both simple and compound, was covered by De Morgan in his *Elements of Arithmetic* [**12**, pp. 150–60], but his tracts extended this treatment to include the rudiments of acturial mathematics, introducing the students to the complex calculation of annuities based on mortality tables. Permutations and combinations were also covered, leading directly to elementary problems in probability theory. Interestingly, for this subject, De Morgan relied on a popular algebra primer written by a former student, Isaac Todhunter, [**35**] who by the 1850s had become a successful textbook author.

Additional geometrical topics were also begun in this class, albeit at a fairly introductory level, the first being algebraic geometry. At this stage, problems set primarily involved either tracing curves or finding the intersection of two lines by solving simultaneous equations. Their initiation into projective geometry reached a slightly higher stage, proceeding as far as Pascal's and Brianchon's theorems. They and much more besides, would be repeated in full in his lower senior class.

3.3 The Lower Senior Class. By the time they entered the lower division of De Morgan's senior class, the majority of his students would have completed at least one year of mathematical study. That year would have provided the students with a study programme of considerable intensity. However, this pales in comparison with the level of material covered during the following year, as illustrated by the number of relevant notebooks still in existence: in comparison to the 80 such documents relating to the junior classes, there are no fewer than 247 notebooks concerning material covered by the two divisions of the senior class.

According to published sources at least, the lower senior course began with an introduction to spherical trigonometry. Again, De Morgan's tracts on this topic supplemented both his lectures and a book on the subject—in this case, a small textbook he had written in 1834. His tracts included further explanation and examples of various points, including statements and proofs of the standard formulae for spherical triangles, and problems such as finding areas, inscribing and circumscribing circles, and supplemental triangles.

Compared to just five items on spherical trigonometry, the number of individual notebooks containing material relating to conic sections is well over twenty; moreover, De Morgan's treatment often varies from tract to tract. To begin with, the conics would have been defined purely geometrically. De Morgan would then introduce the closely-related topic of projective geometry, although, judging from the higher junior tracts, his students would already have received some introduc-

tion to the subject by this time. De Morgan's justification was that "the method of projections establishes the more general and more difficult properties of the conic sections with greater ease than the ordinary methods". [**49**, f. 1] His projective geometry largely consisted of an analysis of various properties and peculiarities of projective figures, such as colinearity and involution, with all demonstrations relying on neatly drawn diagrams and Euclidean-style proofs.

Once the class had reached a certain level of proficiency in projective geometry, De Morgan would employ algebraic geometry to give alternative demonstrations of similar—and, in some cases, the same—results. Having already defined straight lines and circles algebraically in the higher junior class, he began this level with a discussion of the general second degree equation $ay^2 + bxy + cx^2 + dy + ex + f = 0$, and considered the curves generated by its different variations. In such a way, he was able to give yet another introduction to the conic sections, extending the treatment to include algebraic treatments of results originally proved using projective geometry.[3]

At this stage the class would have reached a fairly advanced level of algebra; indeed, by this time, their algebraic exercises included multiplying and dividing polynomials, and solving cubics using Cardano's and Ferrari's methods. Among other algorithms taught by De Morgan in the theory of equations was Horner's method,[4] a procedure for approximating roots of equations with no exact solution. He later described his motivation for introducing this method, and the results his students obtained after applying it to the equation $x^3 - 2x = 5$:

> In 1831, Fourier's posthumous work on equations [**18**, pp. 209–17] showed 33 figures of solution, got with enormous labour. Thinking this is a good opportunity to illustrate the superiority of the method of W. G. Horner, not yet known in France, and not much known in England, I proposed to one of my classes, in 1841, to beat Fourier on this point, as a Christmas exercise. I received several answers, agreeing with each other, to 50 places of decimals. In 1848, I repeated the proposal, requesting that 50 places might be exceeded: I obtained answers of 75, 65, 63, 58, 57, and 52 places. [**5**, p. 292]

It is here that we begin to detect a new feature in De Morgan's teaching: a desire to acquaint the more advanced pupils with recent mathematical developments. Tract #25, for example, contains material that, while ostensibly concerned with the theory of equations, would nowadays be considered as part of complex analysis, being straight from the pages of recent works by Cauchy and Argand. It also includes several new proofs of the existence of a root of every equation, including a paper by De Morgan on the subject, pasted in the back as usual, although his advice to students was: "Read Argand first, and then examine Cauchy's". [**43**, f. 17]

The lower seniors would also have been presented with many of the latest results in modern analysis, especially in their study of infinite series. But it was only after they had been given a thorough grounding in algebraic and analytic operations, especially regarding the meaning and significance of limits, that they were initiated into the subject of the differential calculus. From the tracts, we can

[3] Much of De Morgan's treatment of conic sections in his tracts on algebraic geometry was taken directly from George Salmon's *Treatise on Conic Sections*, Hodges and Smith, Dublin, 1847.

[4] Named after William George Horner (1786–1837), a school-teacher from southwest England, due to a paper he published in 1819.

be sure that a fair number of students encountered problems; a particular difficulty concerned the differential coefficient of a function—a problem still encountered by students today.

> A beginner sees $(1+x^2)^3$, and remembering that x^3 gave $3x^2$, he writes down $3(1+x^2)^2$. He ought to have written $3(1+x^2)^2 \times 2x$. The truth is that he has correctly answered *a question*,—but not *the question which was asked.* [**45**, f. 17]

De Morgan's initial teaching of integration proceeded no further than finding areas under curves. However, there is evidence that he began elementary instruction on differential equations in this class, although this involved little more than defining basic notions such as the order of an equation, the integrating factor, and how to find general and singular solutions. Such an introduction would have been of little use to those who chose to end their mathematical studies at this point. But these final subjects were to provide a background for the detailed course of study reserved for students who proceeded to De Morgan's higher senior class.

3.4 The Higher Senior Class. Attending De Morgan's lectures as far as his lower senior class would have enabled the average student to pass the B.A. examination at the University of London, as well as to move on to the study of natural philosophy in the college. However, for those exceptionally capable (and keen) students who perhaps wished to try for an M.A. degree, it was advisable to enter University College's highest mathematical class. This course was obviously the most technically demanding and, although the class would never have been huge, was one to which De Morgan clearly devoted much time and attention.

This is evinced by the 138 notebooks he wrote for this class, more than for any other division of his students. This high number of tracts is explained by the fact that fewer, if any, elementary textbooks were available on the topics of his higher senior lectures. For much of this section of the course, the most useful work would have been his *Differential and Integral Calculus*, since the subject dominated the material covered. Other areas were also treated, such as further theory of equations, three-dimensional geometry, and probability theory, but their study was vastly outweighed by the amount of time devoted to calculus-related topics.

Chief among these was the study of differential equations, briefly introduced in the lower senior class. As with all of De Morgan's tracts on subjects of some complexity, those dealing with the first principles cover each aspect in careful detail. It is quite obvious from the sheer number of notebooks relating to the various types of equation (around thirty) that De Morgan was anxious that his student's should obtain as much experience and practice of solving them as possible. He even wrote an entire tract containing model solutions to questions on the subject from University of London examination papers. The chief application of differential equations in De Morgan's higher senior tracts was to the study of curves and surfaces, where the subject matter is almost entirely based on the differential geometry contained in Gauss's *Disquisitiones Generales circa Superficies Curvas* of 1827.

The class was also introduced to a second form of differentiation in order to facilitate the subsequent study of mechanics. This was the calculus of variations. Much of the material contained in the tracts is also presented in his *Differential and Integral Calculus* [**11**, pp. 446–75], such as the famous brachistochrone problem of finding "the curve of shortest descent from one curve to another, a heavy point descending upon the curve (supposed hard) by the action of gravity, with no

velocity at the commencement". [**44**, f. 4] But his treatment of the subject, while thorough, was not exhaustive; for example, he directed the more advanced students to "the Memoir of Poisson on the Calculus of Variations, in the twelfth volume of the Memoirs of the Institute". [**11**, p. 454]

In addition to the study of these 'pure' mathematical subjects, De Morgan also managed to include a few items of applied mathematics. Indeed, more time was spent on mathematical applications in the higher senior class than in any other—although the overall proportion was still minute. One subject considered was probability theory, which the students had studied—in its pure form—in the higher junior class. De Morgan would now introduce them to its applications, most notably its use in error theory, a precursor of what would now be called mathematical statistics. De Morgan's teaching of this subject was also heavily influenced by the work of Gauss a few decades before. This is hardly surprising since the main topics in this area, such as the weight of observations and the method of least squares, were all introduced by Gauss. Thus once again, De Morgan can be seen to be acquainting his students with (fairly) recent work on a new and rapidly growing area of mathematical research.

Less recent—but certainly still applied—mathematics is contained in two notebooks on the subject of dynamics. Strictly speaking, this would have been taught by the professor of natural philosophy, but De Morgan's treatment was entirely mathematical, dealing purely with theoretical problems involving the derivation of equations of motion for particles travelling under certain conditions. Moreover, throughout these tracts, he is at pains to stress the distinction between the abstract mathematical notions of velocity and acceleration on the one hand, and the physical phenomena (e.g., force, pressure, and attraction) that cause them. Thus, for example:

> When, as is usual in books on mechanics, *acceleration* is much confounded with *force measured by the acceleration it produces*... —called *accelerating force*—the *centrifugal acceleration*, a law of space, gets the name of *centrifugal force*, whether there be such a force in action or not. [**52**, f. 16]

His motivation for thus trespassing on materials within the domain of mathematical physics was his belief that "the want of sufficient attention to this distinction puts some difficulties in the way of beginners in dynamics". [**52**, f. 1] In other words, he thought that if his students received an adequate notion of velocity and acceleration independently of any physical consideration of the properties of matter, they would be better equipped to understand the subject of dynamics when they came to study natural philosophy.

Having been given a thorough grounding in most areas of contemporary mathematical science, even proceeding far enough in certain subjects to have become acquainted with several aspects of recent research, the student would have several options for further study. Although the concept of a graduate research student did not exist in Britain at this time, the higher senior class would almost certainly have served as a good starting point for those aspiring for an academic career in mathematics, since it not only provided guidance for those aiming for mathematical honours, but also those trying for a London M.A. or preparing to embark on a course of study at Cambridge.

3.5 Overview. In fact, viewing the surviving mathematical tracts as representative of De Morgan's entire syllabus, one is impressed not just by the level to which

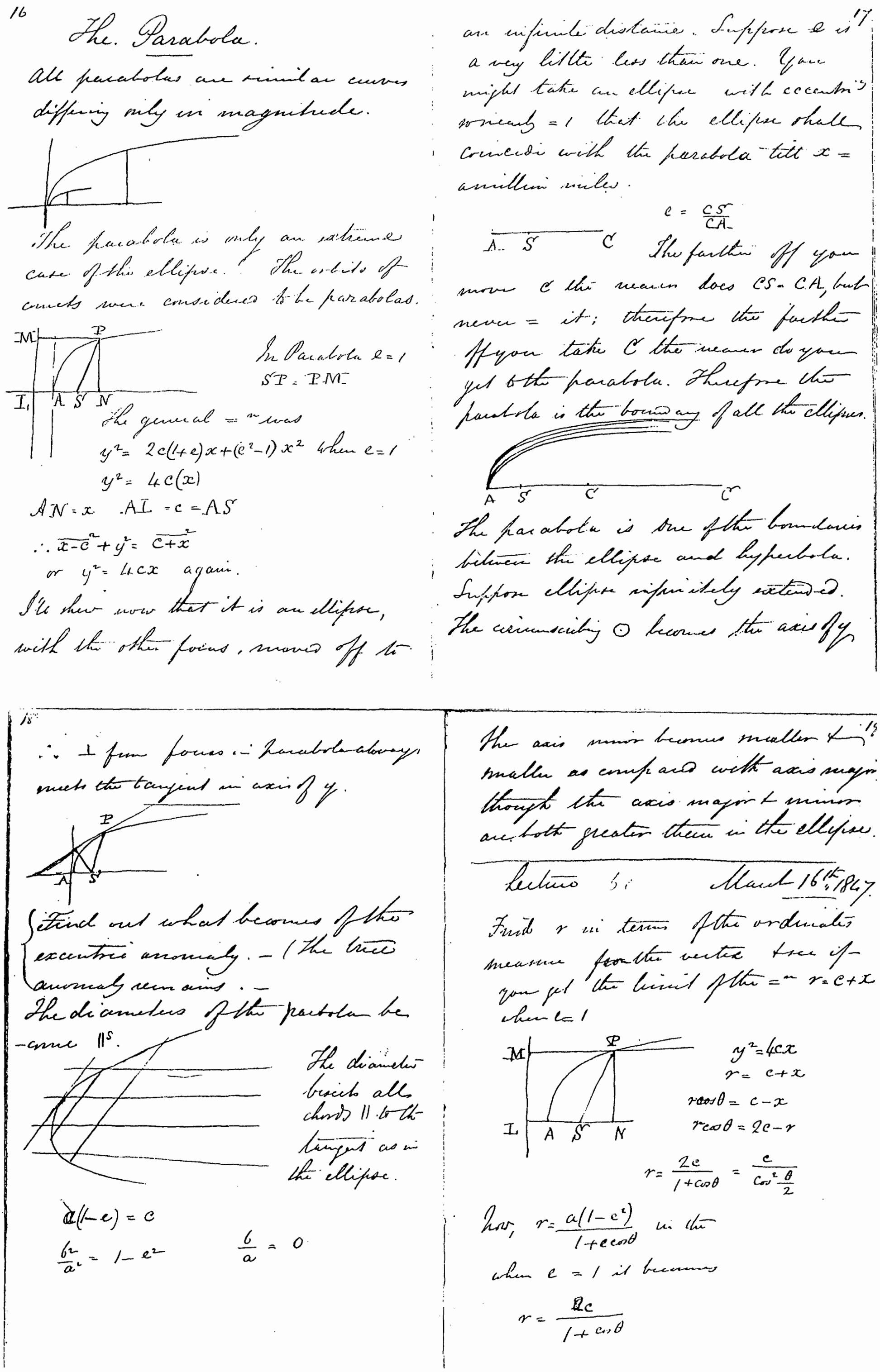
mathematics was taught, but also by the range of topics to which the students were exposed by their professor. To be sure, there is nothing unusual in his basic course structure, whereby the subject is developed from arithmetic and Euclid through the standard branches of algebra, geometry, trigonometry, and calculus; but this is hardly an original feature. Rather, it is the additional, less prominent topics, absent from the course outlines and found only in the tracts, which give the course its variety and make it particularly distinctive. The result is a mathematical course of considerable scope and breadth.

De Morgan's course was as advanced as it was varied. Indeed, it extended almost as far as an undergraduate course could at the time, since developments in many branches (especially analysis and algebra) were transforming the subject almost as it was being taught. This is reflected in the fact that the tracts were constantly being updated. A good example is found in notebooks relating to the theory of equations, one of the most rapidly-expanding areas at this time, which yielded new subjects in the form of complex analysis and group theory. Although never taught by De Morgan, many of the results that contributed to this latter development are contained in a tract written in 1855, featuring "selections from what has been recently done in the higher parts of the theory of equations". [**48**, f. 1] This featured substantial extracts from the second edition of Serret's *Cours d'algèbre supérieure* (1854), the third edition of which, published in 1866, was to mark the first appearance of Galois theory in a textbook.

De Morgan's mathematical tracts give us far more information about the style and structure of his teaching than any printed syllabus or exam paper of the time. But ironically, these manuscripts, invaluable though they undoubtedly are, still do not tell us *exactly* what went on in his lecture room. By virtue of the fact that they were written explicitly to *supplement* the students' notes obtained from the Professor's lectures, the tracts can give us only a general impression of what the students would have been taught in person. Fortunately, however, there are three further sources of information that can bring us one step closer to understanding just what it was like to study mathematics under De Morgan.

4. STUDENT-AUTHORED ACCOUNTS. Two of these sources are extracts from the private writings of two eminent students, the journalist and constitutional author Walter Bagehot and the mathematical economist and logician William Stanley Jevons. However, the third is perhaps the most valuable, for two reasons. Firstly, because it was written by a student of more average ability; and secondly, because it is the student's original college notebook, in which he transcribed De Morgan's lectures as they happened. This comparatively academically undistinguished student was one John Golch Hepburn, also destined to achieve no particular eminence following his graduation. However, Hepburn's notebook provides us with a unique insight as to what the student would have experienced in De Morgan's lecture room 150 years ago.

The manuscript contains notes from 21 of De Morgan's lower senior lectures on algebraic geometry and the differential calculus, delivered between 11 March and 13 May 1847. The first begins with a study of the ellipse, considering aspects such as area and conjugate diameters. The next lecture, on 13 March, deals with Kepler's Laws, orbit-time calculations, and an introduction to parabolae. Hyperbolae and asymptotes are treated five days later. By 27 March, the emphasis was on tangents and chords to conic sections. After an absence from two or three lectures, Hepburn's notes resume on 16 April, when sections of cylinders, cones, and



16

The Parabola.

All parabolas are similar curves differing only in magnitude.

The parabola is only an extreme case of the ellipse. The orbits of comets were considered to be parabolas.

In Parabola $e = 1$, $SP = PM$

The general equation was

$y^2 = 2c(1+e)x + (e^2-1)x^2$ when $e=1$

$y^2 = 4c(x)$

$AN = x \quad AI = c = AS$

$\therefore \overline{x-c}^2 + y^2 = \overline{c+x}^2$

or $y^2 = 4cx$ again.

I'll shew now that it is an ellipse, with the other focus, moved off to

17

an infinite distance. Suppose e is a very little less than one. You might take an ellipse with eccentricity $= 1$ that the ellipse shall coincide with the parabola till $x =$ a million miles.

$e = \frac{CS}{CA}$

A S C The farther off you move C the nearer does $CS = CA$, but never $=$ it; therefore the farther off you take C the nearer do you get to the parabola. Therefore the parabola is the boundary of all the ellipses.

The parabola is one of the boundaries between the ellipse and hyperbola. Suppose ellipse infinitely extended. The circumscribing ⊙ becomes the axis of y

18

$\therefore \perp$ from focus in parabola always meets the tangent in axis of y.

Find out what becomes of the excentric anomaly. — (The true anomaly remains. —

The diameters of the parabola become $\parallel^s$.

The diameter bisects all chords $\parallel$ to the tangent as in the ellipse.

$a(1-e) = c$

$\frac{b^2}{a^2} = 1 - e^2 \qquad \frac{b}{a} = 0$

19

The axis minor becomes smaller & smaller as compared with axis major though the axis major & minor are both greater than in the ellipse.

Lecture 5. March 16th 1847

Find r in terms of the ordinate measured from the vertex & see if you get the limit of the equation $r = c + x$ when $e = 1$

$y^2 = 4cx$

$r = c + x$

$r\cos\theta = c - x$

$r\cos\theta = 2c - r$

$r = \frac{2c}{1+\cos\theta} = \frac{c}{\cos^2\frac{\theta}{2}}$

Now, $r = \frac{a(1-e^2)}{1+e\cos\theta}$ in the ellipse

when $e = 1$ it becomes

$r = \frac{2c}{1+\cos\theta}$

Figure 2. John Hepburn's undergraduate notes on De Morgan's Lectures from March 1847.

spheres were under discussion. Less than a week later the students were being introduced to the differential calculus.

De Morgan clearly approached the new subject at some considerable speed since, on its first day, he was teaching derivatives of fundamental expressions and the product and quotient rules, yet, by 24 April, two days later, he was differentiating x^x. By the end of the month, physical notions such as velocity had been introduced, with tangent/normal and max/min problems brought in on 1 May. Maclaurin's theorem was proved for convergent series in the next lecture, followed by Taylor's and Lagrange's theorems, together with related problems. By 8 May, the class had been introduced to the calculus of finite differences, the notes concluding with an introduction to the calculus of operations.

Reference to section 3.3 confirms that Hepburn's notes correspond very closely to topics dealt with in De Morgan's tracts for his lower senior class, but perhaps more remarkable is how rapidly the Professor propelled his students through the subject. His introduction to the calculus took him a little over two weeks, consisting of just seven lectures. In that time, he discussed first principles, including foundational concepts such as limits, as well as derivatives of functions, fundamental rules, and elementary applications, before moving on to some crucial results in analysis. It is little wonder that he provided tracts for his students to augment their lecture notes!

De Morgan's homework problems were numerous and far from trivial. A few examples from Hepburn's lecture notes indicate the standard of De Morgan's homework questions at the lower senior level:

> Determine area of parabola as extreme case of area of ellipse. Suppose axis major become $>$ & $>$; e being nearer & nearer $= 1$. [**38**, f. 25]
>
> Required the [Maclaurin] developments of ε^{ax}, $(1+x)^n$, $\sin x$, $\cos x$, $\tan x$, and $\varepsilon^{\cos x}$ to 8^{th} power at least. [**38**, f. 193]
>
> Try to give a geometrical proof of the ratio of two magnitudes wh$^{\text{h}}$ vanish is the same thing as the ratio of their diff. Coeff$^{\text{ts}}$. [**38**, f. 197]

In addition to such problems, Hepburn's notes are permeated with references to recommended reading. Perhaps the most intriguing citation, contained in the lecture on the foundations of the calculus, was "See *Leipzig Acts* 1684".[5] [**38**, f. 137] Thus it would certainly appear from this text that De Morgan's course was not for the faint-hearted, yet perhaps the only detail absent in the document is any indication of how difficult the student actually found it. For this information we are obliged to refer to our two remaining sources, which fortunately shed some considerable light on this question.

The sources are the diaries and correspondence of Walter Bagehot and Stanley Jevons, who attended De Morgan's lectures during the 1840s and 1850s, respectively. From both it would appear that mathematics under De Morgan was stimulating but never easy. Thus we find Bagehot writing in 1843: "De Morgan has been taking us through a perfect labyrinth lately; he was quite lost by the whole

[5]This refers to the 1684 edition of the journal *Acta Eruditorum Lipsienium*, which contained the first publication by Leibniz on the differential calculus.

class for one lecture, but we are, I hope, getting better, and more gleg[6] at the uptake. We have been discussing the properties of infinite series, which are very perplexing." [**1**, p. 118] And, as he approached his final exams, Bagehot's workload intensified: "I have been reading some of the Theory of Numbers, which De Morgan says is the best exercise for the head possible, and certainly is a hard stretch for my reading powers and memory." [**1**, p. 159]

Stanley Jevons experienced De Morgan's teaching a decade later than Bagehot, and at two different periods. These are recorded in detail in the diary and correspondence written by him during his college years. It is thus in his memoirs that perhaps the fullest and most candid account of experiences as a student of De Morgan can be found. These examples give an evocative description of one man's study of mathematics at University College under the tuition of Augustus De Morgan:

> In mathematics we are just beginning the theory of equations, and during the last week have got through Descartes', Fourier's, and Sturm's theorems of the limits of the roots of equations. They are the most truly difficult things we have come to, and I do not thoroughly understand them yet. [**23**, p. 29]

> ...one learns more and more to adore De Morgan as an unfathomable fund of mathematics. We were delighted the other day when, in the higher senior, he at last appeared conscious that a demonstration about differential equations, which extended through the lecture, was difficult; he promised, indeed, to repeat it. But then one is disappointed to find that the hardest thing he gives in any of his classes is still to him a trifle, and that the bounds of mathematical knowledge are yet out of sight. [**23**, p. 150]

5. CONCLUSION: A COMPARISON OF CONTEMPORARIES. It is not only through contemporary student accounts that we can determine the calibre of De Morgan's teaching. In later years, many of his students recalled the effect he had on them. More than half a century after experiencing his teaching in the late 1840s, the English historian Thomas Hodgkin wrote:

> Towering up intellectually above all his fellows, as I now look back upon him, rises the grand form of the mathematician, Augustus De Morgan, known, I suppose to each succeeding generation of his pupils as 'Gussy'. A stout and tall figure, a stiff rather waddling walk, a high white cravat and stick-up collars in which the square chin is buried, a full but well chiselled face, very short-sighted eyes peering forth through gold-rimmed spectacles; but above all such a superb dome-like forehead, as could only belong to one of the kings of thought: that is my remembrance of De Morgan, and I feel in looking back upon his personality that his is one of the grandest figures that I have known. [**2**, p. 80]

Hodgkin was not the only non-mathematician on whom De Morgan made a substantial impact. Reminiscing in 1921, the lawyer James Bourne Benson affirmed that "De Morgan [was] looked upon with awe" [**3**, f. 3] by the undergraduates of his day. The distinguished chemist Sir Henry Enfield Roscoe went further

[6] A Scots word (found in the poetry of Robert Burns) meaning astute, quick, keen, or alert.

still, opining that De Morgan was more than "merely a mathematician and a unique teacher; he was one of the profoundest and subtlest thinkers of the nineteenth century". [**32**, p. 25] "As a teacher of mathematics," wrote Jevons, "De Morgan was unrivalled. ... De Morgan's writings however excellent, give little idea of the perspicuity and elegance of his *viva voce* expositions, which never failed to fix the attention of all who were worthy of hearing him," [**24**, p. 8]

Jevons was later to acknowledge the profound effect of De Morgan on his intellectual development, and it is clear that the careers of many other former students were also influenced in some way by De Morgan's teaching. Francis Guthrie, the originator of the Four-Colour Conjecture, became a professor of mathematics in South Africa; E. J. Routh moved to Cambridge, becoming one of the most successful mathematical coaches in its history; and Isaac Todhunter achieved renown for a highly successful series of textbooks as well as his research into various aspects of the history of mathematics.

This influence over students reminds one of the effect of Karl Weierstrass on many of those who attended his lectures at the University of Berlin in the latter half of the nineteenth century. While Weierstrass had far more first-rate mathematics students than De Morgan (among them Otto Hölder, Adolf Hurwitz, Felix Klein, Hermann Minkowski, Gösta Mittag-Leffler, Hermann Schwarz, and E. H. Moore), the image of a mathematician who "became a recognised master, primarily through his lectures" [**4**, p. 221] is consistent to both.

One could not provide a greater contrast to the didactic methods of De Morgan and Weierstrass than those of their earlier contemporary Augustin-Louis Cauchy. As a professor at the Ecole Polytechnique in Paris from 1816, Cauchy taught mathematics for trainee engineers. Yet, while his nineteenth-century biographer C. A. Valson extols his virtues as a teacher who "never left a subject until he had completely exhausted and elucidated it", [**53**, I, p. 64] the records of the Ecole contain a damning report of 1821 in which it is claimed that "numerous warnings have been given, for 5 years, to Mr. Cauchy to undertake to simplify his methods and to conform exactly to the programmes [of lectures]." [**19**, II, p. 711] Furthermore, it is asserted that "there has sometimes been ... a lack of clarity in his lectures ... " [**19**, II, p. 712]

Cauchy was also admonished by his superiors for "the lack of that aplomb that one must generally suppose of a *professeur* already celebrated in the sciences". [**19**, II, p. 711] In particular, his punctuality was criticised when he arrived for a lecture "10 minutes after the gathering of the Students in the amphitheatre". [**19**, II, p. 711] This characteristic was certainly not shared with De Morgan, who consistently placed great emphasis on the precision of his arrival. Indeed, he was apparently "so punctual and so regular in the performance of his college duties that his passage to and from his classes served as a time-piece to observant students". [**28**, p. 115] Moreover, during his time at University College, he became one of its most conscientious and respected professors, lecturing from 9 to 10am and 3 to 4pm every day except Sundays. Indeed, according to Roscoe, "the trouble he took with students was extraordinary". [**32**, p. 25]

De Morgan's innate mathematical propensity was enriched by strong communicative skills, which, together with a talent for presenting complex ideas in an intelligible form and a pithy lecturing style, resulted in the ability to captivate his audience irrespective of the topic he was treating. It would also appear from more than one source that, as with other areas of his instruction, in order to foster correct notions in his students, De Morgan's keen sense of humour was often

employed as a pedagogic tool:

> One thing which made his classes lively to men who were up to his mark, was the humorous horror he used to express at our blunders, especially when we took the conventional or book view instead of the logical view. The bland "hush!" with which he would suppress a suggestion which was simply stupid, and the almost grotesque surprise he would feign when a man betrayed that, instead of the classification by logical principles, he was thinking of the old unmeaning classification by rule in the common school-books, were exceedingly humorous, and gave a life to the classes beyond the mere scope of their intellectual interests. [**15**, pp. 97–8]

Here again, De Morgan's formal but good-humoured style contrasts with that of some of his contemporary counterparts. It has been said that "it was only gradually that Weierstrass acquired the masterly skill in lecturing extolled by his later students. Initially his lectures were seldom clear, orderly, or understandable." [**4**, p. 221] Moreover, claimed Felix Klein (himself a highly successful lecturer), the "imposing" personality of the great man "gave his lectures a distinctly uncongenial, authoritarian quality." [**25**, p. 163]

De Morgan was able to balance congeniality with firm discipline to maintain a serious but convivial atmosphere in his lecture room. It was not a skill all of his fellow lecturers were able to achieve. James Clerk Maxwell was arguably one of the foremost British mathematical physicists of the time when he lectured at King's College London in the early 1860s. Yet, remarkable though his scientific credentials may have been, "as a teacher of raw youths, ... he did not prove to be a success ... and it seems not unlikely that the students were too much for him." [**22**, p. 247] Maxwell's inability to maintain order in his classes was exacerbated by additional shortcomings as a lecturer. A recent biographer explains: "The evidence is that, as a teacher, he had unusual difficulties. His delivery was poor. He could control neither the speed of his thoughts nor the flights of his mind .... Very likely only the occasional, particularly brilliant student could follow his lectures." [**36**, p. 100]

The same could be said of the teaching of James Joseph Sylvester, himself a former student of De Morgan, who, while an outstanding pure mathematician, was by no means a clear lecturer. In fact, as De Morgan later recalled. "When he was with us [as professor of natural philosophy at University College from 1837–41] he was an entire failure: whether in lecture room or in private exposition, he could not keep his team of ideas in hand." [**37**] Yet, forty years later, the very qualities that had made him unsuccessful as an undergraduate lecturer proved inspiring to his graduate students at Johns Hopkins University. As one later recalled: "One could not help being inspired by such teaching, and many of us were led to investigate on lines which he touched upon." [**25**, p. 81] Thus, a style of lecturing far less structured and meticulously arranged than that of De Morgan, was no less successful in holding the attention of the able students.

So what can we conclude from these comparisons of teaching characteristics of De Morgan and his contemporaries? What are the distinguishing features of a great teacher? First of all, the lecturer must be capable of delivering tuition by means of a clear and systematically planned series of lectures to form a structured and intelligible course. Clearly, De Morgan was amply equipped to do this, while Sylvester, Cauchy, and even Weierstrass to begin with, were not. Secondly, he must be able to maintain order in the lecture room so as to keep the attention of the

class focused on the subject material. Here again, De Morgan satisfies this criterion, whereas others, most notably Maxwell, do not.

But finally, one must be able to *inspire* one's students with a love and fascination for their subject. De Morgan did this by concentrating on inculcating a deep understanding of fundamental principles rather than a mere skill with techniques and manipulation. Of course, Weierstrass, Klein, and Sylvester achieved success, but this was largely attained by teaching *graduates* whose interest in the subject was already strong. De Morgan's achievement lies in the fact that he was able to persuade so many undergraduates (even those who took the subject no further) of the beauty and allure of mathematics.

Without question Cauchy, Weierstrass, Klein, Maxwell, and Sylvester were all great mathematicians. Some (Weierstrass and Klein in particular) were great teachers as well. But whether De Morgan could also be called a "great" mathematician is debatable. True, he played a major role in the development of symbolic logic and its introduction into mainstream mathematics; he also contributed to the growth of abstract algebra, and promoted the use of Cauchy's limit-based approach to the calculus, but his mathematical output, while by no means trivial, hardly merits the term "great". As a researcher he is clearly not in the same league as those already listed. But, for his work in the lecture room, he certainly deserves a place in the first rank, possessing in abundance all the attributes of a memorable and effective educator. Indeed, if De Morgan's mathematical reputation had to rest on any one achievement, it would have to be as a teacher to whom the term "great" could truly be applied.

ACKNOWLEDGMENTS. The author is grateful to the libraries of the University of London and University College London for their help and assistance with the archival material. He also thanks John Fauvel, Karen Parshall, and Ivor Grattan-Guinness for helpful comments made during the preparation of this article.

REFERENCES

1. E. I. Barrington, *Life of Walter Bagehot*, Longman & Co., London, 1914.
2. H. H. Bellot, *University College London 1826–1926*, University of London Press, London, 1929.
3. J. B. Benson, "Some Recollections of University College in the Sixties", MS (1921), University College Archives, Materials for the history of UCL, Mem. 1B/3.
4. K-R. Biermann, "Karl Weierstrass", in *Dictionary of Scientific Biography*," C. C. Gillispie, ed., 16 vols., New York: Scribner's 1970–80, 14:219–24.
5. A. De Morgan, *A Budget of Paradoxes*, Longmans, Green, and Co., London 1872.
6. A. De Morgan, "An Introductory Lecture delivered at the Opening of the Mathematical Classes in the University of London, Novr. 5th, 1828", University College Archives, MS. ADD.3.
7. A. De Morgan, On indirect demonstration, *Phil. Mag.* (4) **3** (1852) 435–138
8. A. De Morgan, On mathematical instruction, *Quart. J. Education* **1** (1831) 264–79.
9. A. De Morgan, On the method of teaching geometry, part 2, *Quart. J. Education* **6** (1833) 237–51.
10. A. De Morgan, *The Connexion of Number and Magnitude*, Taylor & Walton, London, 1836.
11. A. De Morgan, *The Differential and Integral Calculus*, Baldwin and Cradock, London, 1842.
12. A. De Morgan, *The Elements of Arithmetic*, 5th ed., Taylor, Walton & Maberly, London, 1853.
13. A. De Morgan, *Thoughts suggested by the establishment of the University of London: An introductory lecture, delivered at the opening of the Faculty of Arts, in University College, Oct. 16, 1837*, Taylor & Walton, London, 1837.
14. A. De Morgan, *Trigonometry and Double Algebra*, Taylor, Walton, & Maberly, London, 1849.
15. S. E. De Morgan, *Memoir of Augustus De Morgan*, Longmans, Green, & Co., London, 1882.
16. J. M. Dubbey, The introduction of the differential notation to Great Britain, *Ann. of Sci.* **19** (1963) 37–48.
17. P. C. Enros, The Analytical Society (1812–1813): Precursor of the Renewal of Cambridge Mathematics, *Historia Math.* **10** (1983) 24–47.
18. J-B. J. Fourier, *Analyse des équations déterminées*, Firmin Didot Frères, Paris, 1831.

19. I. Grattan-Guinness, *Convolutions in French Mathematics, 1800–1840*, 3 vols. Birkhäuser Verlag, Basel, 1990.
20. I. Grattan-Guinness, *The development of the foundations of mathematical analysis from Euler to Riemann*, MIT Press, Cambridge, Mass., 1970.
21. N. Guicciardini, *The Development of Newtonian Calculus in Britain 1700–1800*, Cambridge University Press, Cambridge, 1989.
22. F. J. C. Hearnshaw, *The Centenary History of King's College London 1828–1928*, G.G. Harrap & Co., London, 1929.
23. H. A. Jevons ed., *Letters and Journal of W. Stanley Jevons*, Macmillan & Co., London, 1886.
24. W. S. Jevons, De Morgan, *Encyclpædia Brittanica*, 11th ed., Cambridge University Press, Cambridge, 1910, 8:8–10.
25. K. H. Parshall and D. E. Rowe, *The Emergence of the American Mathematical Research Community 1876–1900: J. J. Sylvester, Felix Klein, and E. H. Moore*, American Mathematical Society, Providence, and London Mathematical Society, London, 1994.
26. H. M. Pycior, Augustus De Morgan's Algebraic Work: The Three Stages, *Isis* **74** (1983) 211–226.
27. H. M. Pycior, George Peacock and the British origins of symbolic algebra, *Historia Math.* **8** (1981) 23–45.
28. A. C. Ranyard, Obituary notice of Augustus De Morgan, *Monthly Notices of the Roy Astron. Soc.* **32** (1871–72) 112–118.
29. A. Rice, Augustus De Morgan: historian of science, *Hist. of Sci.* **34** (1996) 201–240.
30. A. Rice, Inspiration or Desperation? Augustus De Morgan's appointment to the Chair of Mathematics at London University in 1828, *British. J. Hist. Sci.* **30** (1997) 257–274.
31. J. L. Richards, Augustus De Morgan, The history of mathematics, and the foundations of algebra, *Isis* **78** (1987) 7–30.
32. H. E. Roscoe, *The Life and Experiences of Sir Henry Enfield Roscoe*, Macmillan & Co., London, 1906.
33. *Second Statement by the Council of the University of London, Explanatory of the Plan of Instruction*, Richard Taylor, London, 1828.
34. G. C. Smith, *The Boole-De Morgan Correspondence 1842–1864*, Clarendon Press, Oxford, 1982.
35. I. Todhunter, *Algebra for the use of colleges and schools*, Macmillan & Co., Cambridge, 1858.
36. I. Tolstoy, *James Clerk Maxwell: A Biography*, Canongate, Edinburgh, 1981.
37. University College London Archives: London Mathematical Society Papers, Letter from Augustus De Morgan to Thomas Archer Hirst, 29 June 1865.
38. University College London Archives, MS.ADD.5, "Lectures on Algebraic Geometry and the Calculus delivered in University College, London, by Prof. A. De Morgan. Session, 1846–1847".
39. *University College, London. Session 1839–40. Faculty of Arts and Laws*, Richard and John E. Taylor, London, 1839.
40. *The University College London Calendar for the Session 1853–4*, Walton & Maberly, London, 1853.
41. *University College, London. Calendar. Session 1866–67*, Walton & Maberly, London, 1866.
42. University of London Library Special Collections, (hereafter ULL), MS. 775/17, "Euclid. Book V".
43. ULL, MS.775/55, "On the roots of Algebraic Expressions".
44. ULL, MS.775/186, "Calculus of Variations II".
45. ULL, MS.775/197, "Rules of Differentiation".
46. ULL, MS.775/202, "On some points of Geometrical Demonstration".
47. ULL, MS.775/203, "On the Two Forms of the Universal Affirmative Proposition".
48. ULL, MS.775/244, "Elimination &c".
49. ULL, MS.775/261, "Elements of projection and projective properties".
50. ULL, MS.775/271, "Distinction of Arithmetic and Algebra".
51. ULL, MS.775/284, "On Limits".
52. ULL, MS.775/297, "Examples on Velocity and Acceleration".
53. C. A. Valson, *La vie et les travaux du baron Cauchy*, 2 vols., Gauthier-Villars, Paris, 1868.

ADRIAN RICE received his B.Sc. in mathematics from University College London in 1992. In 1997 he was awarded a Ph.D. in the history of mathematics by Middlesex University. He is now a visiting assistant professor at the University of Virginia. His principal area of research is mathematics and mathematical education in nineteenth-century Britain, with especial reference to London and, in particular, the work of Augustus De Morgan. He has written several papers on Victorian mathematics.
University of Virginia, Charlottesville, VA 22903-3199
ar6n@virginia.edu

The TEAM Approach to Investing

Frank Gerth III

1. DESCRIPTION OF STRATEGIES. We consider an investor with a portfolio of "stock-like" investments (e.g., an S&P 500 Index fund) and "cash-like" investments (e.g., Treasury bills, CDs, and high-grade commercial paper). $S(t)$ and $C(t)$ denote the values at time t of this "stock fund" and "cash fund," respectively, and

$$V(t) = S(t) + C(t) \tag{1.1}$$

denotes the total portfolio value at time t. The ratios $r(t,h) = [V(t+h) - V(t)]/V(t)$ have been studied exhaustively for many such portfolios and for time increments h ranging from days, weeks, and months, to quarters, years, and even longer. For short intervals h, the ratios $r(t,h)$ may be positive or negative, are seldom large, and are difficult to distinguish from independent identically distributed random variables with constant (or slowly varying) means and variances.

At an initial time t_0, we suppose amounts S_0 and C_0 are invested in the stock fund and cash fund, respectively. We compare the results of two strategies. The first is a buy-and-hold strategy with no exchanges between the stock fund and cash fund. The second involves periodically reallocating money between the stock fund and cash fund in a way that maximizes a certain function that is described in the next section. We call the first the BH (Buy-and-Hold) strategy and the second the TEAM (Target Equity Allocation Management) strategy. Our main result (Proposition 1) is that the TEAM strategy produces a higher expected total portfolio value than does the BH strategy for the same level of risk.

What are the reallocations in the TEAM strategy? Pick a convenient time interval h, and let $t_i = t_0 + ih$ for $i = 1, 2, \dots, n$. Suppose the investor moves money back and forth between the stock fund and cash fund only at the instants t_i for $i = 1, 2, \dots, n-1$, in an attempt to obtain a large total portfolio value $V(t_n) = S(t_n) + C(t_n)$. For $i = 1, 2, \dots, n$, let s_i (respectively, c_i) denote the rate of return of the stock fund (respectively, cash fund) during the i-th time period. In the TEAM strategy, money is moved between the stock fund and cash fund at the instants t_i so that

$$S(t_i) = S_0(1 + c_1)(1 + c_2) \cdots (1 + c_i) \tag{1.2}$$

for $i = 1, 2, \dots, n-1$. The reason for these stock fund allocations is explained in the next section. In contrast, in the BH strategy,

$$S(t_i) = S_0(1 + s_1)(1 + s_2) \cdots (1 + s_i) \tag{1.3}$$

for each i.

2. ANALYSIS OF STRATEGIES. We suppose that the s_i's (respectively, c_i's) are independent identically distributed random variables with means μ_s and variances σ_s^2 (respectively, means μ_c and variances σ_c^2). We assume $\mu_s > \mu_c$ and $\sigma_s^2 > \sigma_c^2$. The first inequality reflects the fact that stock-like assets tend to appreciate more rapidly than cash-like assets, while the second inequality expresses the fact that

stock-like assets appreciate at a riskier (less predictable) rate than cash-like assets. Generally speaking, investors are paid to bear risk, with extra pay for extra risk.

Let Δ_i denote the dollar amount transferred from the stock fund to the cash fund at time t_i. If $\Delta_i < 0$, money is transferred from the cash fund to the stock fund. (In the BH strategy $\Delta_1 = \cdots = \Delta_{n-1} = 0$.) Write $S_i = S(t_0 + ih)$, $C_i = C(t_0 + ih)$, and $V_i = V(t_0 + ih)$. For any sequence $\Delta_1, \ldots, \Delta_{n-1}$,

$$S_i = S_{i-1}(1 + s_i) - \Delta_i \tag{2.1}$$

and

$$C_i = C_{i-1}(1 + c_i) + \Delta_i \tag{2.2}$$

for $i = 1, \ldots, n$. We may suppose $\Delta_n = 0$ since Δ_n is of no consequence. Then

$$V_i = S_i + C_i = V_{i-1}(1 + c_i) + S_{i-1}(s_i - c_i) \tag{2.3}$$

for each i. Note that $V(t) = S(t) + C(t)$ is continuous across the boundaries (instants) separating consecutive time intervals, even though $S(t)$ and $C(t)$ are not.

The investor may cause the quantities $S_0, S_1, \ldots, S_{n-1}$ to assume any values in the intervals $[0, V_0], [0, V_1], \ldots, [0, V_{n-1}]$ by choosing S_0 and $\Delta_1, \ldots, \Delta_{n-1}$ appropriately. Even these constraints may be eliminated by "selling short" stock-like assets or cash-like assets, e.g., borrowing money in the latter case. Therefore, it seems natural to ignore the constraints $0 \le S_i \le V_i$ during the initial search for an effective investment strategy.

Now let $g_0 = 1$ and

$$g_i = g_{i-1}(1 + c_i) = (1 + c_1) \cdots (1 + c_i)$$

for $i \ge 1$. Making the substitutions $V_i = g_i X_i$ and $S_i = g_i u_{i+1}$ in (2.3) gives

$$X_i = X_{i-1} + R_i u_i \tag{2.4}$$

where

$$R_i = (s_i - c_i)/(1 + c_i) \tag{2.5}$$

for each i. Then $X_0 = V_0$ and

$$X_i = X_0 + R_1 u_1 + \cdots + R_i u_i \tag{2.6}$$

for each $i \ge 1$. The investor's choice of the numbers $u_1, \ldots, u_n$ is entirely unrestricted, and the investor presumably chooses $u_1, \ldots, u_n$ in an attempt to maximize an appropriate function of portfolio performance.

For subsequent calculations it is useful to assume that $\sigma_c^2 = \mathrm{Var}(c_i) = 0$ for each i, so that $c_1 = \cdots = c_n = \mu_c = c$, a numerical constant, and then $g_i = (1 + c)^i$ for each i. In practice, $\mathrm{Var}(c_i)$ is usually so much smaller than $\mathrm{Var}(s_i)$ that it seems natural to ignore $\mathrm{Var}(c_i)$. Indeed, the availability of a "riskless rate c of return" is an essential feature of the "Black-Scholes environment" in which most financial analysis is performed.

Next let $M_n = V_0 g_n$, which is the amount the cash fund would contain if the entire initial investment V_0 had been placed in the cash fund and left alone. We would like the terminal value V_n of our chosen portfolio of stocks and cash to exceed M_n by a large amount. However, to take into account the riskiness of our portfolio, we choose to maximize the function

$$f(V_n - M_n) = E(V_n - M_n)/\sqrt{\mathrm{Var}(V_n - M_n)}\,, \tag{2.7}$$

which increases with the expected value of $V_n - M_n$ and decreases with the standard deviation of $V_n - M_n$. This function is essentially the Sharpe reward-to-variability ratio [**2**]. Now from (2.6)

$$\begin{aligned} V_n - M_n &= g_n(X_n - X_0) = g_n(R_1 u_1 + \cdots + R_n u_n) \\ &= (1+c)^n (R_1 u_1 + \cdots + R_n u_n) \end{aligned} \tag{2.8}$$

under our assumption that $c_i = c$ for each i. Then (2.5) and our assumption that the s_i's are independent identically distributed random variables imply that the R_i's are independent identically distributed random variables with means $(\mu_s - c)/(1+c)$ and variances $\sigma_s^2/(1+c)^2$.

We now consider a special case for the u_i's; namely, we assume that the u_i's are chosen in a way that does not depend on the stock fund rates of return $s_1, s_2, \ldots, s_n$. Then (2.7) and (2.8) imply

$$f(V_n - M_n) = K\langle \mathbf{e}, \mathbf{u}\rangle / \|\mathbf{u}\|_2 \leq K\|\mathbf{e}\|_2 = K\sqrt{n} \tag{2.9}$$

where $\mathbf{u}$ is the vector $(u_1, \ldots, u_n)$, $\mathbf{e}$ is the vector with n components each equal to 1, and $K = (\mu_s - c)/\sigma_s$. The inequality in (2.9) is a consequence of the Cauchy-Schwarz inequality, with equality if and only if $\mathbf{u} = \lambda\mathbf{e}$ for some positive scalar λ. Since $u_1 = g_0 u_1 = S_0$, then equality in (2.9) occurs precisely when $\mathbf{u} = S_0\mathbf{e}$ with S_0 positive. Then

$$S_i = g_i u_{i+1} = S_0(1+c)^i \tag{2.10}$$

for $i = 1, \ldots, n-1$. Note that (2.10) is the same as (1.2) under the assumption that $c_i = c$ for each i. This optimal strategy corresponds to the choices $\Delta_i = (s_i - c)S_{i-1}$ for $i = 1, \ldots, n-1$ in (2.1). Each $S_i/S_{i-1} = 1 + c$, which means that the stock fund allocations increase at the riskless rate c. This strategy is called the Target Equity Allocation Management (TEAM) strategy since it allocates resources between the riskier stock fund and more conservative cash fund in a way calculated to achieve the modest equity targets $S_i = S_0(1+c)^i$ in the stock fund, while transferring anticipated surpluses to the cash fund.

Although the TEAM strategy maximizes the function (2.7) among all strategies for which the u_i's do not depend on the stock fund rates of return $s_1, \ldots, s_n$, we should expect some feedback type strategies (in which the u_i's depend on the s_j's) to produce greater values for $f(V_n - M_n)$. In the buy-and-hold (BH) strategy, the u_i's do depend on the s_j's since $\Delta_1 = \cdots = \Delta_{n-1} = 0$ imply that

$$(1+c)^i u_{i+1} = S_i = S_0(1+s_1)\cdots(1+s_i) \tag{2.11}$$

for $i = 1, \ldots, n-1$. Hence there is still some work to do to show that the TEAM strategy imparts a higher value to the function $f(V_n - M_n)$ than does the BH strategy.

To avoid confusion with other strategies, let B_n rather than V_n denote the terminal portfolio value for the BH strategy. From (2.5) and (2.11), R_i depends on s_i, whereas u_i depends on $s_1, \ldots, s_{i-1}$. Thus, R_i and u_i are independent, and (2.8) implies

$$E(B_n - M_n) = (1+c)^n\left(\bar{R}_1\bar{u}_1 + \cdots + \bar{R}_n\bar{u}_n\right) \tag{2.12}$$

where

$$\bar{R}_i = E(R_i) = (\mu_s - c)/(1+c) \tag{2.13}$$

$$\bar{u}_i = E(u_i) = S_0(1+\mu_s)^{i-1}/(1+c)^{i-1} \tag{2.14}$$

for $i = 1, \ldots, n$. A straightforward but lengthy calculation shows that

$$\operatorname{Var}(B_n - M_n) > (1 + c)^{2(n-1)} \sigma_s^2 \left(\bar{u}_1^2 + \cdots + \bar{u}_n^2\right). \tag{2.15}$$

Then (2.7), (2.12), (2.13), and (2.15) imply

$$f(B_n - M_n) < K\langle \mathbf{e}, \mathbf{w} \rangle / \|\mathbf{w}\|_2 < K\|\mathbf{e}\|_2 = K\sqrt{n} \tag{2.16}$$

where $\mathbf{w} = (\bar{u}_1, \ldots, \bar{u}_n)$, $\mathbf{e} = (1, \ldots, 1)$, and $K = (\mu_s - c)/\sigma_s$. The second inequality in (2.16) is a consequence of the Cauchy-Schwarz inequality and the fact that $\mathbf{w} \neq \lambda \mathbf{e}$ for a scalar λ since $\mu_s > c$ in (2.14).

Now if T_n is the terminal portfolio value for the TEAM strategy, then

$$f(B_n - M_n) < K\sqrt{n} = f(T_n - M_n). \tag{2.17}$$

For the TEAM strategy, the analogs of (2.12) and (2.15) are

$$E(T_n - M_n) = (1 + c)^n \left(\bar{R}_1 S_0 + \cdots + \bar{R}_n S_0\right) = (1 + c)^{n-1} n(\mu_s - c) S_0 \tag{2.18}$$

$$\operatorname{Var}(T_n - M_n) = (1 + c)^{2(n-1)} n \sigma_s^2 S_0^2. \tag{2.19}$$

Since $\mu_s > c$, then (2.12), (2.14), (2.15), (2.18), and (2.19) imply that for the same initial allocations in the BH strategy and in the TEAM strategy, the BH strategy has a higher expected terminal portfolio value and a higher variance than does the TEAM strategy. Next, observe that the value $f(T_n - M_n) = K\sqrt{n}$ in (2.17) does not depend on the initial allocations in the stock fund and cash fund. Hence we could increase the initial allocation to an amount S_0' in the stock fund in the TEAM strategy (while correspondingly decreasing the initial allocation to an amount C_0' in the cash fund) until $\operatorname{Var}(T_n' - M_n) = \operatorname{Var}(B_n - M_n)$, where T_n' is the TEAM terminal portfolio value for initial investments of S_0' and C_0' in the stock and cash funds, respectively. Then (2.7) and (2.17) imply $E(T_n') > E(B_n)$.

We now list the primary assumptions that are used in our derivation and then state the proposition we have proved.

The cash fund rate of return is constant over all time periods. (2.20)

The stock fund rates of return in each time period are independent identically distributed random variables. (2.21)

The expected stock fund rate of return exceeds the cash fund rate of return. (2.22)

There are no taxes or transaction costs. (2.23)

Proposition 1. *Suppose an amount S_0 (respectively, S_0') is invested initially in the stock fund in the BH strategy (respectively, TEAM strategy) and an amount C_0 (respectively, C_0') is invested initially in the cash fund in the BH strategy (respectively, TEAM strategy). Let B_n (respectively, T_n') denote the terminal portfolio value for the BH (respectively, TEAM) strategy for these initial allocations. Suppose S_0' and C_0' are chosen so that $(S_0' + C_0') = (S_0 + C_0)$ and so that the standard deviations of T_n' and B_n are equal. Then under the assumptions* (2.20) *through* (2.23), *the expected value of T_n' is greater than that of B_n. Hence for the same level of risk (as measured by standard deviation of terminal portfolio value), the TEAM strategy produces a higher expected terminal portfolio value than does the BH strategy.*

If S_0 is large relative to C_0, then it could happen that $S_0' > (S_0 + C_0)$, in which case $C_0' < 0$. This corresponds to borrowing money in the TEAM strategy. Also, reallocations in the TEAM strategy could require borrowing money.

The assumptions (2.20) through (2.23) are not precisely satisfied in practice, and our model is a very simplified financial model. Nevertheless it is interesting to compare simulation results using the BH and TEAM strategies.

3. SIMULATION RESULTS. We examine the results of some simulations using historical data from 1926 to 1995. Most of the data comes from Ibbotson and Sinquefield [**1**, pp. 54–55]. For rates of return for our stock fund, we use rates of return for Common Stocks (which correspond to the S & P 500 Index); for our cash fund, we use the rates of return for U.S. Treasury Bills. We exclude transaction costs and taxes, and consider 14 non-overlapping 5-year periods from 1926 to 1995. In each 5-year period, the TEAM strategy reallocations occur annually, so $n = 5$. There are no reallocations in the BH strategy during each 5-year period. The data in [**1**] show that the serial correlation for annual rates of return is near zero for a stock fund such as an S & P 500 Index fund, which is consistent with what we would expect from (2.21).

Since the TEAM strategy involves lower risk than the BH strategy when the BH and TEAM strategies have the same initial allocations, one can increase the initial percentage allocation to the stock fund when using the TEAM strategy. Let S_0 (respectively, S_0') denote the initial percentage allocation in the stock fund when using the BH strategy (respectively, TEAM strategy), and set

$$a = S_0'/S_0. \tag{3.1}$$

The TEAM simulation results in Table 1 correspond to $a = 1.4$. We also performed simulations with $a = 1.0$, 1.1, 1.2, 1.3, 1.5, 1.6, 1.7, and 1.8. As one would expect, the average terminal value for the TEAM strategy (and its standard deviation) increased as a increased. However the standard deviation for the TEAM strategy was less than the standard deviation for the BH strategy for $a \leq 1.6$. For $1.2 \leq a \leq 1.6$, the TEAM strategy produced a higher terminal value than the BH strategy in at least 12 of the 14 five-year periods.

TABLE 1 Total Returns for Two Strategies

Strategy	Buy and Hold (BH)		TEAM
Initial Value	$1.0000		$1.0000
Stock fraction	.5000		.7000
Cash fraction	.5000		.3000
Time Periods	Terminal Values ($)		Difference (TEAM − BH)
1926–1930	1.3476	1.5202	.1726
1931–1935	1.0964	1.3220	.2256
1936–1940	1.0150	1.1413	.1263
1941–1945	1.6010	1.6461	.0451
1946–1950	1.3221	1.3977	.0756
1951–1955	1.9973	1.9555	−.0418
1956–1960	1.3336	1.4382	.1046
1961–1965	1.5135	1.5951	.0816
1966–1970	1.2413	1.2481	.0068
1971–1975	1.2477	1.3398	.0921
1976–1980	1.6874	1.7785	.0911
1981–1985	1.8094	1.9154	.1060
1986–1990	1.6226	1.7063	.0837
1991–1995	1.6933	1.7713	.0780
Average	1.4663	1.5554	.0891
Standard Deviation	.2823	.2520	.0643

For $a = 1.4$ in Table 1, the average difference of .0891, i.e., an average 5-year total return difference of 8.91% for the TEAM strategy over the BH strategy, is significantly greater than zero at the 1% level for the t-statistic for the variable (TEAM − BH) terminal value. One might argue that this variable is not normally distributed, and hence the t-statistic might not be appropriate. Further statistical justification for concluding that the (TEAM − BH) terminal value is significantly greater than zero is the fact that the TEAM terminal value exceeded the BH terminal value in 13 of the 14 time periods. Hence with an appropriate choice of a, the TEAM strategy can produce a higher terminal value than the BH strategy, with no greater risk than the BH strategy.

The data in Table 1 correspond to the assumption that each stock fund investment S_i' in the TEAM strategy is limited to the total value T_i'. If borrowing is allowed, the TEAM strategy results are slightly better.

REFERENCES

1. R. Ibbotson and R. Sinquefield, *Stocks, Bonds, Bills, and Inflation: Historical Returns* (1926–1987), Financial Analysts Research Foundation, Charlottesville, VA, 1989.
2. W. Sharpe, Mutual fund performance, *J. Business* **39** (1966) 119–138.

FRANK GERTH III is Professor of Mathematics at the University of Texas at Austin. He received a B.A. degree from Rice University and a Ph.D. degree from Princeton University under the supervision of Kenkichi Iwasawa. He was an Instructor at the University of Pennsylvania and has spent a semester at the Mathematical Sciences Research Institute in Berkeley, CA. His primary research area is number theory.
The University of Texas at Austin, Austin, TX 78712
gerth@math.utexas.edu

From the MONTHLY 25 Years Ago...

THE DERIVATIVE SONG

Words by Tom Lehrer—Tune: "There'll be Some Changes Made"

You take a function of x and you call it y,
Take any x-nought that you care to try,
You make a little change and call it delta x,
The corresponding change in y is what you find nex',
And then you take the quotient and now carefully
Send delta x to zero, and I think you'll see
That what the limit gives us, if our work all checks,
Is what we call dy/dx,
It's just dy/dx

MONTHLY 81 (1974) 490

A Physically Motivated Further Note on the Mean Value Theorem for Integrals

William J. Schwind, Jun Ji, and Daniel E. Koditschek

The purpose of this Note is to extend the following result by Zhang concerning the mean value theorem for integrals [**4**], which in turn was an extension of Jacobson's result [**1**].

Theorem 1. *Suppose the function f is continuous on the interval $[a, b]$, and is k times differentiable at a with $f^{(i)}(a) = 0$ $(i = 1, 2, \ldots, k-1)$, $f^{(k)}(a) \neq 0$. If ξ_x is such that*

$$\int_a^x f(t)\,dt = f(\xi_x)(x-a), \tag{1}$$

then

$$\lim_{x \to a} \frac{\xi_x - a}{x - a} = \frac{1}{(k+1)^{\frac{1}{k}}}. \tag{2}$$

As a side note, Jacobson [**1**] states that his result may fall into the category of "interesting facts we once knew, but have now forgotten." This indeed seems to be the case, as the right-hand side of (2) is derived in [**2**, p. 78].

We extend Theorem 1 to a considerably larger function class, which includes functions such as $f(t) = \sqrt{t-a}$ and $f(t) = 1/\sqrt{t-a}$ that do not satisfy the hypotheses of Theorem 1. Additionally, we apply the new results to obtain approximations of integrals appearing in familiar engineering settings. In fact our consideration of this problem was motivated by the desire to approximate, in closed form, integrals arising from a class of central force problems and to do so independently of the particular mathematical form of the force law itself.

Theorem 2. *Suppose f is continuous on $(a, b]$ and g is integrable on (a, b) with $g(t) \geq 0$ for all $t \in (a, b)$. Let $x \in (a, b]$. If both $\lim_{t \to a}(f(t) - K)/(t-a)^r$ and $\lim_{t \to a} g(t)/(t-a)^s$ exist and are nonzero for some constant K, some nonzero r, and some $s > -1$ with $r + s > -1$, then*

1. *there exists a $\xi_x \in (a, x]$ such that*

$$\int_a^x f(t)g(t)\,dt = f(\xi_x)\int_a^x g(t)\,dt; \tag{3}$$

2. *for any such choice of* ξ_x,

$$\lim_{x\to a}\frac{\xi_x - a}{x-a} = \left(\frac{s+1}{r+s+1}\right)^{\frac{1}{r}}. \tag{4}$$

Proof: Define $C_1 = \lim_{t\to a}(f(t)-K)/(t-a)^r$, $C_2 = \lim_{t\to a} g(t)/(t-a)^s$, $\epsilon_1(t) = (f(t)-K)/(t-a)^r - C_1$, and $\epsilon_2(t) = g(t)/(t-a)^s - C_2$, so that

$$f(t) = K + C_1(t-a)^r + \epsilon_1(t)(t-a)^r \tag{5}$$

and

$$g(t) = C_2(t-a)^s + \epsilon_2(t)(t-a)^s, \tag{6}$$

where $\epsilon_i(t) \to 0$ as $t \to a$ for $i = 1,2$.

First we must justify that the integral on the left hand side of (3) actually exists. Fix $\delta \in (a,b)$ such that $\epsilon_i(t) \le 1$ for $a < t < \delta$. Now $|f(t)g(t)| \le K_1(t-a)^s + K_2(t-a)^{s+r}$ for some constants K_1 and K_2. The Dominated Convergence Theorem [**3**, p. 291] ensures that $\int_a^\delta f(t)g(t)\,dt$ exists. Therefore, $\int_a^x f(t)g(t)\,dt$ exists.

Proof of (1). Case 1 ($r > 0$). It is easily seen from (5) that $\lim_{t\to a} f(t) = K$. Define $F(a) = K$ and $F(t) = f(t)$ for $t \in (a,b]$. Clearly, $F(t)$ is continuous on $[a,b]$. By applying the Integral Mean Value Theorem [**3**, p. 281] to $F(t)$, there exists a ξ_x in (a,x) such that

$$\int_a^x f(t)g(t)\,dt = \int_a^x F(t)g(t)\,dt = F(\xi_x)\int_a^x g(t)\,dt = f(\xi_x)\int_a^x g(t)\,dt.$$

Case 2 ($r < 0$). Without loss of generality take $C_1 > 0$ and $\int_a^x g(t)\,dt > 0$. Then (5) requires that $\lim_{x\to a} f(t) = +\infty$. Thus there exists a positive δ such that

$$f(t) > \frac{\int_a^x f(t)g(t)\,dt}{\int_a^x g(t)dt} \equiv \eta, \quad \text{for all } a < t < a+\delta.$$

Therefore, a ξ_x satisfying (3), if it exists, is not contained in $(a, a+\delta)$. For $a_1 \in (a, a+\delta)$, f is continuous on $[a_1, x]$ and $f(t) > \eta$ on $(a, a_1]$. Hence,

$$\min_{t\in(a,x]} f(t) = \min_{t\in[a_1,x]} f(t).$$

Therefore,

$$\max_{t\in[a_1,x]} f(t) \ge f(a_1) > \eta \ge \min_{t\in(a,x]} f(t) = \min_{t\in[a_1,x]} f(t).$$

By applying the Intermediate Value Theorem to $f(t)$ on $[a_1, x]$, we obtain $\xi_x \in [a_1, x] \subseteq (a,x]$ such that $f(\xi_x) = \eta$.

Proof of (2). Substituting (5) and (6) into the left hand side of (3), one sees easily that

$$\begin{aligned}\int_a^x f(t)g(t)\,dt = {} & \frac{C_2K(x-a)^{s+1}}{s+1} + \frac{C_1C_2(x-a)^{r+s+1}}{r+s+1} \\ & + C_2\int_a^x \epsilon_1(t)(t-a)^{r+s}\,dt + K\int_a^x \epsilon_2(t)(t-a)^s\,dt \\ & + C_1\int_a^x \epsilon_2(t)(t-a)^{r+s}\,dt + \int_a^x \epsilon_1(t)\epsilon_2(t)(t-a)^{r+s}\,dt. \end{aligned} \tag{7}$$

On the other hand, substituting (5) and (6) into the right hand side of (3), we find

$$
\begin{aligned}
f(\xi_x)\int_a^x g(t)\,dt = {} & \frac{C_2K(x-a)^{s+1}}{s+1} + \frac{C_1C_2(\xi_x-a)^r(x-a)^{s+1}}{s+1} \\
& + \frac{C_2\epsilon_1(\xi_x)(\xi_x-a)^r(x-a)^{s+1}}{s+1} + K\int_a^x \epsilon_2(t)(t-a)^s\,dt \\
& + C_1(\xi_x-a)^r\int_a^x \epsilon_2(t)(t-a)^s\,dt \\
& + \epsilon_1(\xi_x)(\xi_x-a)^r\int_a^x \epsilon_2(t)(t-a)^s\,dt. \qquad (8)
\end{aligned}
$$

Equating (7) and (8) and simplifying gives

$$U\frac{(\xi_x-a)^r}{(x-a)^r} = V, \qquad (9)$$

where

$$
\begin{aligned}
U = 1 + \frac{\epsilon_1(\xi_x)}{C_1} & + \frac{(s+1)\epsilon_1(\xi_x)\int_a^x \epsilon_2(t)(t-a)^s\,dt}{C_1C_2(x-a)^{s+1}} \\
& + \frac{(s+1)\int_a^x \epsilon_2(t)(t-a)^s\,dt}{C_2(x-a)^{s+1}}
\end{aligned}
$$

and

$$
\begin{aligned}
V = \frac{s+1}{r+s+1} & + \frac{(s+1)\int_a^x \epsilon_1(t)(t-a)^{r+s}\,dt}{C_1(x-a)^{r+s+1}} \\
& + \frac{(s+1)\int_a^x \epsilon_2(t)(t-a)^{r+s}\,dt}{C_2(x-a)^{r+s+1}} + \frac{(s+1)\int_a^x \epsilon_1(t)\epsilon_2(t)(t-a)^{r+s}\,dt}{C_1C_2(x-a)^{r+s+1}}.
\end{aligned}
$$

Because $|\xi_x - a| \le |x-a|$ and

$$\lim_{x\to a}\frac{\int_a^x d(t)(t-a)^m\,dt}{(x-a)^{m+1}} = 0 \quad \text{if} \quad \lim_{x\to a} d(t) = 0, \quad m > -1,$$

we obtain

$$\lim_{x\to a} U = 1 \quad \text{and} \quad \lim_{x\to a} V = \frac{s+1}{r+s+1} > 0. \qquad (10)$$

Now (4) follows from (9) and (10). ■

Often, a good choice for K in Theorem 2 is to take $K = \lim_{t\to a} f(t)$ if the limit exists or $K = 0$ otherwise.

A few immediate observations can be made when $g(t) \equiv 1$. In this case (3) is identical to (1), the statement of the Mean Value Theorem used by Jacobson and Zhang.

Observation 1. *If* $g(t) \equiv 1$ *in Theorem* 2, *then* $s = 0$ *and*

$$\lim_{x \to a} \frac{\xi_x - a}{x - a} = \left(\frac{1}{r + 1} \right)^{\frac{1}{r}} \tag{11}$$

for some nonzero $r > -1$.

Notice that the form of the limiting value in (11) is identical to that in (2), but $r \in (-1, \infty) \setminus \{0\}$, while k is only a natural number.

Observation 2. *Theorem* 1 *is a special case of Theorem* 2.

To see this, assume the hypothesis in Theorem 1; then

$$\lim_{t \to a} \frac{f(t) - f(a)}{(t - a)^k} = \lim_{t \to a} \frac{f^{(k-1)}(t)}{k!(t - a)} = \lim_{t \to a} \frac{f^{(k-1)}(t) - f^{(k-1)}(a)}{k!(t - a)} = \frac{f^{(k)}(a)}{k!} \neq 0,$$

where the first equality is obtained by using L'Hospital's rule $k - 1$ times while the the third one follows from the definition of $f^{(k)}(a)$.

Observation 3. *If, motivated by* (4), *we approximate* ξ_x *by*

$$\hat{\xi}_x = a + \left(\frac{s + 1}{r + s + 1} \right)^{\frac{1}{r}} (x - a) \quad \textit{for } x \textit{ near } a,$$

and replace ξ_x *by* $\hat{\xi}_x$ *in* (3), *we obtain an approximation scheme to the integral*

$$\int_a^x f(t) g(t)\, dt \approx f(\hat{\xi}_x) \int_a^x g(t)\, dt \quad \textit{for } x \textit{ near } a. \tag{12}$$

A CENTRAL FORCE EXAMPLE. Consider the simple central force problem in which a mass on a spring is restricted to move in the vertical direction. Let the spring potential be given by $U(y)$, where y is the distance from the ground to the mass. Then the dynamics are given by

$$\ddot{y} = -g - \frac{DU(y)}{m}, \tag{13}$$

where g is the acceleration due to gravity and $-DU(y)$ is the spring force; as a matter of notation, $DU(y) = U'(y)$.

Since we assume no losses, the total energy is a constant of motion and we can formulate the integral for time as

$$T(y) = \frac{1}{\sqrt{2g}} \int_{y_i}^{y} \frac{d\psi}{\sqrt{(y_b - \psi) + \frac{1}{mg}(U(y_b) - U(\psi))}}, \tag{14}$$

where y_b is location of the mass when the vertical velocity is zero.

Under reasonable assumptions on the spring potential, which will be made clear in the following, we can approximate (14) using Observation 3.

Other problems, having integrals that resemble (14), could also be considered. Examples are the integrals for swing time and swing angle of a rotating mass on a spring or the integral for swing time of a simple pendulum. It is well known that the solution of this latter example can be formulated as an elliptic integral and thus several well-known approximation techniques can be applied. The results

presented in this Note could be used as well, but additionally, and more importantly, they may also be applied if there are other forces, such as a torsional spring, acting on the pendulum.

We focus on applying the results of Theorem 2 to our central force problem. Suppose we desire to approximate the length of time it takes for the mass to move from the point of maximal compression, y_b, to some other location, y. The result is given by (14) with $y_i = y_b$. Furthermore, suppose we wish to solve this problem without assuming a particular functional form for the spring potential, $U(y)$. In such a case, it is impossible to integrate (14) in closed form. However, (12) provides an approximation to the integral of interest.

To apply Theorem 2, we need to factor the integrand of (14) into the product of two functions, f and g. This factorization is by no means unique and each choice results in a slightly different approximation to the integral. We illustrate two possible choices and discuss the advantages and disadvantages of each.

Let us first consider

$$g_1(y) = 1 \quad \text{and} \quad f_1(y) = \frac{1}{\sqrt{(y_b - y) + \frac{1}{mg}(U(y_b) - U(y))}}.$$

In this case $s_1 = 0$. Choosing $K = 0$ and under reasonable assumptions on $U(y)$, that is,

$$\lim_{y \to y_b} \frac{U(y) - U(y_b)}{y - y_b} = U'(y_b) \quad \text{exists and is not equal to } -mg, \tag{15}$$

we find $r_1 = -1/2$. Note that the exception $U'(y_b) = -mg$, implies $\ddot{y}_b = 0$. Since by definition $\dot{y}_b = 0$, this corresponds to an equilibrium of the system. From (12), we have an approximation to the integral (14)

$$\hat{T}_1(y) = \frac{y - y_b}{\sqrt{2g}\sqrt{\left(y_b - \hat{\xi}_y^1\right) + \frac{1}{mg}\left(U(y_b) - U\left(\hat{\xi}_y^1\right)\right)}}, \tag{16}$$

where $\hat{\xi}_y^1 = y_b + (1/4)(y - y_b)$.

Now consider

$$g_2(y) = \frac{1}{\sqrt{y - y_b}} \quad \text{and} \quad f_2(y) = \frac{1}{\sqrt{-1 + \frac{1}{mg}\left(\frac{U(y_b) - U(y)}{y - y_b}\right)}}.$$

Here, we find $s_2 = -1/2$. If we assume (15), we can choose $K = \lim_{y \to y_b} f_2(y)$. In order to solve for r_2, we need to apply L'Hospital's theorem. This, however, does not provide enough information to determine r_2—we need to know more about the structure of $U(y)$. If, for example, we assume

$$\lim_{y \to y_b} \frac{U'(y)(y - y_b) - (U(y) - U(y_b))}{(y - y_b)^2} = U''(y_b) \neq 0 \quad \text{exists}, \tag{17}$$

we find $r_2 = 1$. If, however, $U''(y_b) = 0$, we must apply L'Hospital again and we find that if

$$\lim_{y \to y_b} \frac{U''(y)(y - y_b)^2 - (U'(y)(y - y_b) - (U(y) - U(y_b)))}{(y - y_b)^3}$$
$$= U'''(y_b) \neq 0 \quad \text{exists},$$

then $r_2 = 2$.

Let us assume (17) holds; this implies $r_2 = 1$. In this case, using (12), we have another approximation to the integral (14),

$$\hat{T}_2(y) = \frac{1}{\sqrt{2g}\sqrt{-1+\frac{1}{mg}\left(\frac{U(y_b)-U(\hat{\xi}_y^2)}{\hat{\xi}_y^2-y_b}\right)}}\int_{y_b}^{y}\frac{1}{\sqrt{\psi-y_b}}d\psi$$

$$= \frac{2\sqrt{y-y_b}}{\sqrt{2g}\sqrt{-1+\frac{1}{mg}\left(\frac{U(y_b)-U(\hat{\xi}_y^2)}{\hat{\xi}_y^2-y_b}\right)}}, \tag{18}$$

where $\hat{\xi}_y^2 = y_b + (1/3)(y - y_b)$.

In this approximation strategy, each approach has its own advantages and disadvantages. The second approach has the advantage of extracting the "dominant" behavior of g and integrating that exactly, but has the drawback of requiring more explicit knowledge of the spring potential law in order to calculate r. The first approach, while not offering the exactness of the second, allows greater flexibility because its approximation is based only upon the "uninformed" factorization ($g(t) \equiv 1$), which allows r to be determined with only minimal knowledge of the spring potential law. Therefore, one's choice of approximation depends on the application of interest.

In this Note we provide two different approximations to (14). From Observation 3, we know that these approximations are good for y close to y_b; however, our proof shows that these approximations may be suitable over larger intervals.

ACKNOWLEDGMENTS. We thank Professors Charles Doering and Charles Kicey for their helpful suggestions regarding this paper. This work was supported in part by National Science Foundation Grant IRI – 9612357.

REFERENCES

1. Bernard Jacobson, On the Mean Value Theorem for Integrals, *Amer. Math. Monthly* **89** (1982) 300–301.
2. Ganesh Prasad, *Six Lectures on the Mean Value Theorem of the Differential Calculus*, University of Calcutta, Calcutta, 1931.
3. Karl R. Stromberg, *An Introduction to Classical Real Analysis*, Wadsworth, Inc., Belmont, CA, 1981.
4. Zhang Bao-lin, A Note on the Mean Value Theorem for Integrals, *Amer. Math. Monthly* **104** (1997) 561–562.

Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109
{schwindw, kod}@eecs.umich.edu

Department of Mathematics and Computer Science, Valdosta State University, Valdosta, GA 31698
junji@valdosta.edu

Hopping Hoops Don't Hop

James P. Butler

The dynamical behavior of a massless hoop with an attached point mass under the influence of gravity is an old chestnut, with some surprising features, including the question of whether or not it can "hop". It appears in Littlewood's *Miscellany* **[1]** and most recently in this MONTHLY **[2]**. That there are many interesting aspects of this problem was suggested to me many years ago by the late Prof. J.C. Miller of Pomona College, who may have gotten a hint of the trickier parts directly from Littlewood himself. The purpose of this Note is to show that Littlewood's and Tokeida's asserted solution is wrong, even with a more realistic hoop and more realistic friction, and to suggest some approaches to a self consistent investigation of the rolling hoop.

In its simplest form, the problem asks for the behavior of a rough massless hoop of radius R with a point mass M attached to its rim, rolling in a vertical plane on a level floor under gravity. Implicitly in Littlewood's problem, and explicitly in this Note, we understand the concept of massless to be the limit of a positive hoop mass tending to 0; otherwise the rotational behavior of the hoop in free fall is undefined. The idea of "rough" is central to the problem; it is (roughly) defined as a no-slip constraint at the point of contact of the hoop and the floor. Let $\vartheta \in [0,\pi]$ be the angle from the radius vector to the mass, measured from vertical; other ranges of ϑ are not considered in this Note. If the total energy is equal to the gravitational potential energy of the mass one diameter above the floor, the "solution" to the problem is the assertion by Littlewood that the hoop "lifts off the ground" or by Tokeida that the mass "pulls the hoop up" at $\vartheta = \pi/2$; this particular value of ϑ depends on the total energy, but is especially simple in the case cited. Denote the normal force conferred by the floor on the hoop by n. The hopping conclusion is alleged to follow from the observation that, for zero kinetic energy at $\vartheta = 0$ and for the rim constrained to be in no-slip contact with the floor, $n > 0$ for $\vartheta < \pi/2$ and $n < 0$ for $\vartheta > \pi/2$. The evaluation of $n(\vartheta)$ is elementary using Newton's second law and conservation of energy (the rough condition is conservative). An equivalent argument **[2]** is that the motion of the point mass either follows a cycloid for $0 \le \vartheta \le \pi/2$, or its free fall parabolic preference for $\pi/2 \le \vartheta$; here the upper limit was not specified.

This solution is wrong on both mathematical and physical grounds. To show this, we begin with the equations of motion, with the following notational conventions. Nondimensionalize the problem by measuring distance in units of radius R, mass in units of M, and time in units of $\sqrt{R/g}$, where g is gravitational acceleration. In these units, gravitational acceleration is -1. The horizontal and vertical coordinates of the point mass are $x = \vartheta + \sin\vartheta$ and $y = 1 + \cos\vartheta$, respectively. The kinetic energy is given by $(1/2)(\dot{x}^2 + \dot{y}^2) = \dot{\vartheta}^2(1 + \cos\vartheta)$, where the overdot denotes the time derivative, and the potential energy by $y = 1 + \cos\vartheta$. With this notation, Newton's law takes the form

$$\ddot{y} = n - 1 = -\cos(\theta)\dot{\theta}^2 - \sin(\theta)\ddot{\theta}.$$

If we take the total energy to be 2 (the potential energy at $\vartheta = 0$), the conservation of energy yields

$$\dot{\vartheta}^2(1 + \cos\vartheta) = 1 - \cos\vartheta.$$

From these two equations, it is easily verified that n is positive for $\vartheta < \pi/2$, zero for $\vartheta = \pi/2$, and negative for $\vartheta > \pi/2$; since the floor cannot exert a negative normal force on the hoop, one concludes (incorrectly) that the hoop hops past $\vartheta = \pi/2$.

Theorem. *Littlewood's hoop doesn't hop.*

Proof: Let the coordinates of the hoop's center be x_c, y_c. For the hoop to hop, y_c must be greater than one (radius) above the floor (as long as the hoop is in contact with the floor, $y_c = y - \cos\vartheta = 1$). The proof that this cannot happen is by contradiction. Assume that past $\vartheta = \vartheta_0 = \pi/2$ the hoop hops, implying that the mass falls freely, then solve for y_c, and show that it cannot be greater than one. Let t_0 be the time at which $\vartheta = \vartheta_0$ and $n = 0$. Assume that for $t_0 < t < t_1$, the hoop hops, meaning $y_c > 1$. In this interval, the point mass is in free fall; the equations of motion are different from those with rough contact and a positive normal force. In particular, the vertical coordinate of the mass is given by $y = y_0 + \dot{y}_0(t - t_0) - (1/2)(t - t_0)^2$ and ϑ is given by uniform angular rotation $\vartheta = \vartheta_0 + \dot{\vartheta}_0(t - t_0)$, where the subscript 0 indicates the quantity evaluated at $t = t_0^-$. During the hop, therefore,

$$\dot{y}_c = \dot{y} + \sin(\vartheta)\dot{\vartheta} = \dot{y}_0 - (t - t_0) + \sin(\vartheta)\dot{\vartheta}_0,$$

$$\ddot{y}_c = \ddot{y} + \cos(\vartheta)\dot{\vartheta}^2 + \sin(\vartheta)\ddot{\vartheta} = -1 + \cos(\vartheta)\dot{\vartheta}_0^2.$$

At the beginning of the hop, $y_c(t_0) = 1$, $\dot{y}_c(t_0) = 0$, and importantly, $\ddot{y}_c(t_0) = -1$. It follows that for the open time interval $t_0 < t < t_1$, $\ddot{y}_c < 0$. Over this interval therefore, the velocity of the center of the hoop is negative, i.e., it must anti-hop, or push through the floor. This is inconsistent with the assumption of a hop. ■

What this theorem shows is that Littlewood's problem is singular in the sense that there is no solution past t_0 consistent with Newton's laws, an impenetrable floor ($y_c \geq 1$), and no-slip ($x_c = \vartheta$) when $n = 0$. What then does the real hoop do? If we retain Newton's laws and an impenetrable floor, then the no-slip condition must be violated when $n = 0$, implying that the hoop skids. One must check that the skidding solution is self consistent; that this is true is sketched in the following arguments.

There are several issues and questions raised by Littlewood's problem. First, with respect to the hopping conclusion, there is the explicit error in not using the different equations of motion for the different periods. Equivalently, the idea of simultaneously using a constraint such that $y_c = 1$ together with $n < 0$ to argue that $y_c > 1$ is clearly inconsistent. Second, the concept of "rough" contact is not well defined. As we have argued, we may retain a no-slip condition for two objects in contact with $n > 0$, but not for $n = 0$; in what follows we define no-slip conditions only for $n > 0$. Third, what then happens to such a real hoop during the skid phase? Can it subsequently hop? Fourth, if we impose realistic frictional conditions rather than rough ones, can such a hoop hop without skidding, or can it hop following a skid phase?

It is not the purpose of this Note to give complete answers to these questions, but a few points can be made. The idea of a point mass and a massless hoop

suggests a second type of singular character to the equations of motion. This singularity manifests itself as a reduction of the number of equations of motion from three to two when the moment of inertia about the center of mass is zero, and as a concomitant reduction in the number of degrees of freedom for the frictional force from two to one (the vector friction force must point in the direction of the point mass). In approximating the behavior of real hoops, it is thus appropriate to consider distributing the mass on a real hoop so that its center of mass is at a radial distance $\lambda < 1$ and its moment of inertia about that point is $I > 0$; we must investigate the hoop's behavior for λ near 1 and for I near zero. With coordinates x, y generalized to the center of mass ($x = x_c + \lambda \sin \vartheta$, $y = y_c + \lambda \cos \vartheta$) and denoting the tangential frictional force of the floor on the hoop by f, the equations of motion, valid irrespective of no-slip, skidding, or free-fall conditions, are

$$\ddot{x} = f,$$
$$\ddot{y} = n - 1,$$
$$I\ddot{\vartheta} = n(x - x_c) - fy = n\lambda \sin \vartheta - f(1 + \lambda \cos \vartheta).$$

The no-slip condition remains $x_c = \vartheta$. It turns out that in this case as well, our Theorem remains true. The proof by contradiction is similar, and may be sketched as follows. When the hoop hops, we have $n(t_0) = 0$, $\ddot{y}(t_0) = -1$, $\ddot{y}_c(t_0^-) = -1 + \lambda \cos(\vartheta_0)\dot{\vartheta}_0^2 + \lambda \sin(\vartheta_0)\ddot{\vartheta}_0 = 0$, and $\ddot{y}_c(t_0^+) = -1 + \lambda \cos(\vartheta_0)\dot{\vartheta}_0^2$. Eliminating $\ddot{\vartheta}$ from the equations of motion (written in terms of ϑ) easily shows that $f < 0$ and $\ddot{\vartheta} > 0$ when $n = 0$. Comparison of $\ddot{y}_c(t_0^-)$ with $\ddot{y}_c(t_0^+)$ then shows $\ddot{y}_c(t_0^+) < 0$, which as before implies a contradictory anti-hop and therefore a skid. That the skid phase is well defined is shown by the existence of a finite f consistent with keeping the hoop on the floor.

Having established that (at least for $\vartheta \in [0, \pi]$) an ideal hoop and its real cousin must skid if the normal force goes to zero, we now ask for a little more reality in the friction law. This might be Coulomb friction, with skidding occurring at some critical ratio of the tangential and normal force. One might even include a distinction between the ratio at which skidding begins (static friction), and the ratio during the skid (dynamic friction). In any case, the physics is no longer conservative, and the absence of an energy integral means the equations are more difficult to analyze. Nevertheless, some general conclusions may still be drawn. (A) Is the skid forward ($\dot{\vartheta} > \dot{x}_c$: hoop is rotating faster than necessary for no-slip (an accelerating train with slipping wheels)), or backwards ($\dot{\vartheta} < \dot{x}_c$: hoop is rotating slower than required for no-slip (a decelerating train with slipping wheels))? The answer is that both behaviors are possible. Starting from ϑ near zero, the initial tangential force is clearly positive, and if the hoop is sufficiently greased, f/n can match the static friction; the hoop skids forward. If it's not, there is necessarily a backwards skid with $f < 0$. The sign of f at the beginning of the skid must therefore be considered. (B) While skidding, is a hop possible at some t_1? There are several possibilities here. (B1) Call a hop a "smooth take-off" if it occurs when $n = 0$. During the skid $|f/n|$ is fixed for Coulomb friction, and thus a smooth take-off requires $f = 0$ also. In this case (by analogy with the sketched proof), $\ddot{\vartheta}(t_1^-) = 0$, $\ddot{y}_c(t_1^-) = \ddot{y}_c(t_1^+) = 0$, but $\dddot{y}_c(t_1^+) = -\lambda \sin(\vartheta_1)\dot{\vartheta}_1^3 < 0$, again implying an anti-hop. (B2) Call a hop a "semi-smooth take-off" if it occurs when the skidding hoop catches up to the floor, and dynamic friction switches over to static friction. That this additional complication doesn't change matters is easily seen by the observation that a change in force to a different, but finite, value for an infinitesimal amount of time cannot change the subsequent dynamics. (B3) What remains is

the possibility of a "jump hop", associated with impulsive forces, approximated by delta functions in "stick-slip" dynamical problems such as chattering chalk on a blackboard. This phenomenon appears to be the most likely candidate for the origin of the real hop of real hoops, but raises more complicated questions about the physics, and makes the analysis commensurately more difficult.

This analysis of the hopping hoop leads to several conclusions. First, and most important, is that Newton's laws and the kinematical constraint for "rough" contact are in general inconsistent when the normal force is zero. Second, real hoops that hop must skid first, and the subsequent hop cannot be smooth nor semi-smooth. Third, there is a rich structure in the behavior of real hoops: vary λ and I, vary the initial conditions, let ϑ be unbounded, follow the bounce(s) after the hop. Finally, with respect to this isolated singularity in Mr. Littlewood's *Miscellany*, he did say that in practice, "the hoop skids", but seemed to imply this to be due to a realistic friction law rather than a necessary consequence even with an unbounded coefficient of static friction. The answer to his query whether the behavior of the hoop is intuitive is given by the following

Theorem. *The behavior of hopping hoops is not intuitive.*

Proof: By inspection.

REFERENCES

1. J. E. Littlewood, *Littlewood's Miscellany*, Cambridge University Press., Cambridge, 1986.
2. T. F. Tokieda, The Hopping Hoop, *Amer. Math. Monthly* **104** (1997) 152–154.

Physiology Program, Harvard School of Public Health, 665 Huntington Ave., Boston, MA 02115
jbutler@hsph.harvard.edu

Approximation of Hölder Continuous Functions by Bernstein Polynomials

Peter Mathé

In a recent MONTHLY [**5**], a special instance of the Weierstraß approximation theorem attracted attention: approximation of real Lipschitz functions on $[0, 1]$ by Bernstein polynomials

$$B_n(f, x) := \sum_{j=0}^{n} \binom{n}{j} x^j (1-x)^{n-j} f\left(\frac{j}{n}\right).$$

The authors of [**5**] provided a rate of uniform convergence of $B_n(f, \cdot)$ to f using large deviations techniques. It is the aim of this note to discuss the optimal rate of approximation with some historical remarks. More generally we consider the class $\mathrm{Lip}_\alpha(L)$ of Hölder continuous functions with exponent α for some $0 < \alpha \leq 1$ and constant L, i.e., functions that obey

$$|f(x) - f(y)| \leq L|x-y|^\alpha \quad \text{for all } x, y \in [0,1].$$

We present two simple proofs of the following

Theorem 1. *If the function* $f:[0,1]\to\mathbb{R}$ *is Hölder continuous with exponent* α *and constant* L *then*

$$|f(x)-B_n(f,x)|\le L\left(\frac{x(1-x)}{n}\right)^{\alpha/2}\quad \textit{for all } n\in\mathbb{N} \textit{ and all } x\in[0,1]. \tag{1}$$

In contrast to Gzyl and Palacios [**5**], who refine the original Bernstein argument to treat the specialization to Lipschitz functions, we use more direct arguments that are elementary and improve the rate to the optimal one.

The history of this result is worth recounting. Bernstein first used his eponymous polynomials to prove the Weierstraß approximation theorem in 1912 [**2**]. It took more than twenty years before results concerning the rate of convergence of $B_n(f,\cdot)$ to f appeared, by Popoviciu [**9**] and by Kac ([**6**] and [**7**]). While Popoviciu established speed of convergence in terms of the modulus of continuity, Kac originally proved exactly Theorem 1. The first proof we present is in the spirit of Kac and exploits the probabilistic meaning of the Bernstein polynomials. The second version uses Korovkin–type arguments [**1**], as in the nice exposition of the Weierstraß approximation theorem in [**10**, Chapter 1.2]. Other approaches are reviewed in Lorentz's monograph [**8**].

The probabilistic nature of the Bernstein polynomials can be recognized when interpreting the weights $\binom{n}{j}x^j(1-x)^{n-j}$ as point probabilities of a binomial distribution with parameters n and x. There are, however, several ways to realize this distribution. To establish a connection with empirical distribution functions, given n, let $u_1,\dots,u_n$ be random variables that are independent and uniformly distributed on $[0,1]$. Consider the random function $S_n:[0,1]\to[0,1]$ defined by

$$S_n(x):=\frac{1}{n}\sum_{j=1}^{n}\chi_{[0,x)}(u_j),\quad x\in[0,1],$$

where $\chi_{[0,x)}$ denotes the characteristic function of the interval $[0,x)$. S_n takes only the values j/n, $j=0,\dots,n$, with probabilities $P(S_n(x)=j/n)=\binom{n}{j}x^j(1-x)^{n-j}$. Thus at any point x the random function value $nS_n(x)$ is binomially distributed and we have expectation $\mathbf{E}S_n(x)=x$ and variance

$$\mathbf{E}(S_n(x)-x)^2=\frac{x(1-x)}{n}. \tag{2}$$

Proof of Theorem 1: probabilistic version. By construction $B_n(f,x)=\mathbf{E}f(S_n(x))$. Using the triangle inequality and Hölder continuity we obtain

$$|f(x)-B_n(f,x)|\le\mathbf{E}|f(x)-f(S_n(x))|\le L\mathbf{E}|x-S_n(x)|^{\alpha}.$$

Apply the Hölder inequality with parameters $2/\alpha$ and $2/(2-\alpha)$, and arrive at

$$|f(x)-B_n(f,x)|\le L\left(\mathbf{E}|x-S_n(x)|^2\right)^{\alpha/2}.$$

Finally, use the variance formula given in (2) to obtain (1). ∎

Proof of Theorem 1: analytic version Here we consider the mapping

$$f\in C([0,1])\to B_n(f,\cdot)\in C([0,1]),$$

which is a bounded linear operator that maps nonnegative functions into nonnegative polynomials (and thus is *positive*). Moreover, letting $p_0(x) = 1$, $p_1(x) = x$, $p_2(x) = x^2$ denote the monomials of degree at most 2, it is easy to check that

$$B_n(p_0, x) = 1, \quad B_n(p_1, x) = x, \quad B_n(p_2, x) = x^2 + x(1-x)/n. \tag{2.5}$$

Denote $f_x(y) := L|x-y|^\alpha$. Thus for any function $f \in \mathrm{Lip}_\alpha(L)$ we have $-f_x(y) \le f(x) - f(y) \le f_x(y)$ for all $y \in [0,1]$. Applying B_n and using its positivity we obtain $|f(x) - B_n(f, y)| \le B_n(f_x, y)$. Hence, for $y = x$ we arrive at $|f(x) - B_n(f,x)| \le B_n(f_x, x)$.

As a substitute for the Hölder inequality the operator B_n obeys

$$B_n(|g|, x) \le \left(B_n(|g|^{2/\alpha}, x)\right)^{\alpha/2} \tag{3}$$

for any function g. Indeed, the geometric–arithmetic–mean inequality

$$\sigma^{1/p}\tau^{1-1/p} \le \frac{1}{p}\sigma + (1-1/p)\tau, \quad p \ge 1, \sigma, \tau \ge 0,$$

with

$$p := \frac{2}{\alpha}, \ \tau := \left(B_n(|g|^{2/\alpha}, x)\right)^{\alpha/2}, \quad \text{and } \sigma := |g(y)|^{2/\alpha}\tau^{1-2/\alpha} \text{ for } y \in [0,1]$$

yields

$$|g(y)| \le \frac{1}{p}|g(y)|^{2/\alpha}\left(B_n(|g|^{2/\alpha}, x)\right)^{\alpha/2-1} + \frac{p-1}{p}\left(B_n(|g|^{2/\alpha}, x)\right)^{\alpha/2}.$$

Using the positivity of B_n we finally arrive at

$$\begin{aligned} B_n(|g|, x) &\le 1/p\left(B_n(|g|^{2/\alpha}, x)\right)^{\alpha/2} + (1-1/p)\left(B_n(|g|^{2/\alpha}, x)\right)^{\alpha/2} \\ &= \left(B_n(|g|^{2/\alpha}, x)\right)^{\alpha/2}, \end{aligned}$$

which proves (3). Thus, for $g := f_x$ we obtain

$$|f(x) - B_n(f,x)| \le \left(B_n(f_x^{2/\alpha}, x)\right)^{\alpha/2}. \tag{3.5}$$

Using the linearity of B_n and its values at p_0, p_1, and p_2 given in (2.5), and observing that $f_x^{2/\alpha}(y) = L^{2/\alpha}(x^2 - 2xy + y^2)$, we conclude that $B_n(f_x^{2/\alpha}, x) = L^{2/\alpha}x(1-x)/n$. Substituting this equality into (3.5) completes the proof. ∎

For completeness we derive the asymptotically exact behavior of the error. The accuracy of the approximation of Hölder continuous functions by Bernstein polynomials is quantified by

$$e_n(L, \alpha) := \sup_{f \in \mathrm{Lip}_\alpha(L)} \sup_{x \in [0,1]} |f(x) - B_n(f,x)|.$$

Theorem 1 can thus be rewritten as $e_n(L,\alpha) \le L(1/(4n))^{\alpha/2}$. This is refined in

Theorem 2. $\lim_{n\to\infty} n^{\alpha/2} e_n(L, \alpha) = L2^{-\alpha/2}\Gamma((\alpha+1)/2)/\sqrt{\pi}$.

Kac proved that the rate of convergence cannot be improved on the class of Hölder continuous functions by establishing an appropriate lower bound for the specific function $f_{1/2} \in \mathrm{Lip}_\alpha(L)$. For this particular choice we have

$$e_n(L, \alpha) \ge \left|f_{1/2}(1/2) - B_n(f_{1/2}, 1/2)\right| = L\mathbf{E}|S_n(1/2) - 1/2|^\alpha. \tag{4}$$

Popoviciu derived the exact asymptotics of the mean absolute deviation $\mathbf{E}|S_n(x) - x|$, i.e., the case $\alpha = 1$, by providing the explicit representation

$$\mathbf{E}|S_n(x) - x| = 2\binom{n-1}{j} x^{j+1}(1-x)^{n-j}, \tag{5}$$

where j is the unique integer for which $x \in [j/n, (j+1)/n)$. This representation has a long history; see **[4]** for a proof and historical details, beginning in 1730, when De Moivre gave a similar explicit formula for the case $x = 1/2$. The derivation of (5) is combinatorial and the analysis of the asymptotic behavior is non-elementary. Our arguments do not rely on the representation (5). Instead we again try to keep close to probabilistic arguments and use the De Moivre–Laplace Theorem on the binomial approximation of the normal distribution.

Our proof of Theorem 2 requires basic knowledge of probability theory and real analysis. For convenience we split the proof of Theorem 2 into two lemmas. As the lower bound in (4) suggests we are concerned with the asymptotic behavior of $n^{\alpha/2}\mathbf{E}|S_n(1/2) - 1/2|^{\alpha}$, and more generally we study the functions $g_n(x) := n^{\alpha/2}\mathbf{E}|S_n(x) - x|^{\alpha}$. Given α we define $K_\alpha := 2^{\alpha/2}\Gamma((\alpha+1)/2)/\sqrt{\pi}$ and $g(x) := (x(1-x))^{\alpha/2}K_\alpha$.

Lemma 1. $\lim_{n\to\infty} g_n(x) = g(x)$ *for every* $x \in [0,1]$.

Moreover the functions g_n belong to $\mathrm{Lip}_{\alpha/2}(1)$ as can be seen from

Lemma 2. $|g_n(x) - g_n(y)| \le |x - y|^{\alpha/2}$ *for all* $n \in \mathbb{N}$ *and all* $x, y \in [0,1]$.

Suppose for a moment that both lemmas have been proved. Pointwise convergence from Lemma 1 and uniform Hölder continuity from Lemma 2 imply uniform convergence as stated in

Proposition 1. $\lim_{n\to\infty}\sup_{x\in[0,1]}|g_n(x) - g(x)| = 0$.

Proof: Our arguments follow the usual proof of the Arzelà–Ascoli Theorem. In fact, for any $m \in \mathbb{N}$ the set $\{j/m, j = 0, \dots, m\}$ is finite so that Lemma 1 provides some n_0 such that

$$|g_n(j/m) - g(j/m)| \le \left(\frac{1}{2m}\right)^{\alpha/2}, \quad j = 0, \dots, m, n \ge n_0.$$

Thus for an arbitrary $x \in [0,1]$ let j_x be chosen such that $|x - j_x/m| \le 1/(2m)$. Then we conclude, using Lemma 2,

$$\begin{aligned}|g_n(x) - g(x)| &\le |g_n(x) - g_n(j_x/m)| + |g_n(j_x/m) - g(j_x/m)| \\ &\quad + |g(j_x/m) - g(x)| \\ &\le 3\left(\frac{1}{2m}\right)^{\alpha/2}\end{aligned}$$

for $n \ge n_0$ and all $x \in [0,1]$. Since m may be chosen arbitrarily large we have established uniform convergence. ■

Now the proof of Theorem 2 follows easily.

Proof of Theorem 2. We infer from (4) that $n^{\alpha/2}e_n(L, \alpha) \geq Lg_n(1/2)$. Thus Lemma 1 yields $\liminf_{n\to\infty} n^{\alpha/2}e_n(L, \alpha) \geq Lg(1/2)$. On the other hand, interchanging $\sup_{f\in \mathrm{Lip}_\alpha(L)}$ and $\sup_{x\in[0,1]}$ in the definition of $e_n(L, \alpha)$ gives

$$\begin{aligned} n^{\alpha/2}e_n(L, \alpha) &= n^{\alpha/2} \sup_{x\in[0,1]} \sup_{f\in\mathrm{Lip}_\alpha(L)} |f(x) - B_n(f, x)| \\ &\leq Ln^{\alpha/2} \sup_{x\in[0,1]} \mathbf{E}|x - S_n(x)|^\alpha = L \sup_{x\in[0,1]} g_n(x). \end{aligned}$$

Since

$$\left| \sup_{x\in[0,1]} g_n(x) - \sup_{x\in[0,1]} g(x) \right| \leq \sup_{x\in[0,1]} |g_n(x) - g(x)|,$$

Proposition 1 implies that $\sup_{x\in[0,1]} g_n(x) \to \sup_{x\in[0,1]} g(x)$ as $n \to \infty$. Thus $\limsup_{n\to\infty} n^{\alpha/2}e_n(L, \alpha) \leq L\sup_{x\in[0,1]} g(x) = Lg(1/2)$. ∎

We turn to proofs of the lemmas.

Proof of Lemma 1. Since the random variable $nS_n(x)$ is binomially distributed for every x, we may apply the De Moivre–Laplace Theorem, which asserts that

$$P\left(a < \frac{nS_n(x) - nx}{\sqrt{nx(1-x)}} < b\right) \to \sqrt{\frac{1}{2\pi}} \int_a^b e^{-u^2/2}\, du$$

for all $a < b$. Especially for all $t > 0$ this yields

$$P\left(\left|\frac{nS_n(x) - nx}{\sqrt{nx(1-x)}}\right| > t\right) \to \sqrt{\frac{2}{\pi}} \int_t^\infty e^{-u^2/2}\, du. \tag{6}$$

Furthermore, we may use Chebyshev's inequality to conclude that

$$P\left(\left|\frac{nS_n(x) - nx}{\sqrt{nx(1-x)}}\right| > t\right) \leq \frac{\mathbf{E}\left|\frac{nS_n(x) - nx}{\sqrt{nx(1-x)}}\right|^2}{t^2} = \frac{1}{t^2},$$

hence

$$P\left(\left|\frac{nS_n(x) - nx}{\sqrt{nx(1-x)}}\right| > t\right) \leq \min\left\{1, \frac{1}{t^2}\right\}. \tag{7}$$

We recall that $g_n(x) = n^{\alpha/2}\mathbf{E}|S_n(x) - x|^\alpha$. Stieltjes integration allows us to compute

$$\begin{aligned} g_n(x) &= \alpha \int_0^\infty t^{\alpha-1} P\left(\sqrt{n}\,|S_n(x) - x| > t\right) dt \\ &= \alpha(x(1-x))^{\alpha/2} \int_0^\infty s^{\alpha-1} P\left(\left|\frac{nS_n(x) - nx}{\sqrt{nx(1-x)}}\right| > s\right) ds. \end{aligned}$$

Since $\int_0^\infty t^{\alpha-1}\min\{1, t^{-2}\}\,dt < \infty$ and we have pointwise convergence as in (6), the Lebesgue Dominated Convergence Theorem implies

$$\begin{aligned}\lim_{n\to\infty} g_n(x) &= \lim_{n\to\infty} \alpha\int_0^\infty t^{\alpha-1}P\big(\sqrt{n}\,|S_n(x) - x| > t\big)\,dt \\ &= (x(1-x))^{\alpha/2}\int_0^\infty s^\alpha\sqrt{\frac{2}{\pi}}\,e^{-s^2/2}\,ds \\ &= (x(1-x))^{\alpha/2}\sqrt{\frac{2}{\pi}}\,2^{(\alpha-1)/2}\int_0^\infty u^{(\alpha+1)/2-1}e^{-u}\,du \\ &= (x(1-x))^{\alpha/2}2^{\alpha/2}\,\frac{\Gamma((\alpha+1)/2)}{\sqrt{\pi}} = g(x). \qquad\blacksquare\end{aligned}$$

It is evident from the proof that K_α is the αth absolute moment of the standard normal distribution, i.e., $K_\alpha = \mathbf{E}|\gamma|^\alpha$, where γ is standard normal. This could also be guessed, since the De Moivre–Laplace Theorem may be viewed as a special instance of the Central Limit Theorem.

Proof of Lemma 2. Since $\big||a|^\alpha - |b|^\alpha\big| \le |a-b|^\alpha$ for all $a, b \in \mathbb{R}$ and $0 < \alpha \le 1$, we have

$$\begin{aligned}|g_n(x) - g_n(y)| &\le n^{\alpha/2}\mathbf{E}|(S_n(x) - x) - (S_n(y) - y)|^\alpha \\ &\le n^{\alpha/2}\big(\mathbf{E}|(S_n(x) - x) - (S_n(y) - y)|^2\big)^{\alpha/2} \\ &= (n\mathrm{Var}(S_n(x) - S_n(y)))^{\alpha/2}.\end{aligned}$$

Now suppose $x < y$ and recall that $S_n(x) - S_n(y) = -\sum_{j=1}^n \frac{1}{n}\chi_{[x,y)}(u_j)$, which is the sum of n independent random variables. It is well known that then the variance of the sum equals the sum of the variances, which implies

$$\begin{aligned}|g_n(x) - g_n(y)| &\le \left(n^2\mathrm{Var}\left(\frac{1}{n}\chi_{[x,y)}(u)\right)\right)^{\alpha/2} = \big(\mathrm{Var}(\chi_{[x,y)}(u))\big)^{\alpha/2} \\ &= ((y-x)(1-(y-x)))^{\alpha/2} \le |y-x|^{\alpha/2} \qquad\blacksquare\end{aligned}$$

We have thus proved the lemmas and hence Theorem 2. We admit that deriving the exact asymptotics of the error is much more elaborate than the simple arguments leading to Theorem 1. The gain in accuracy is less than $\sqrt{2/\pi}$.

To complete our development of properties of the Bernstein polynomials using the probabilistic representation, we add the following result, originally proved by Brown, Elliott, and Paget [**3**].

Proposition 2. *If the function $f : [0,1] \to \mathbb{R}$ is Hölder continuous with exponent α and constant L then so are the corresponding Bernstein polynomials $B_n(f, \cdot)$.*

Proof: Use the triangle inequality, Hölder continuity, the representation of $S_n(x)$ as an empirical distribution function, and finally the Hölder inequality to derive for

$x \leq y$ the estimates

$$|B_n(f,x) - B_n(f,y)| \leq \mathbf{E}|f(S_n(x)) - f(S_n(y))| \leq L\mathbf{E}|S_n(x) - S_n(y)|^{\alpha}$$

$$= L\mathbf{E}\left|\frac{1}{n}\sum_{j=1}^{n} \chi_{[x,y)}(u_j)\right|^{\alpha} \leq L\left(\mathbf{E}\left|\frac{1}{n}\sum_{j=1}^{n} \chi_{[x,y)}(u_j)\right|\right)^{\alpha}$$

$$\leq L|x-y|^{\alpha}. \qquad \blacksquare$$

This proof is brief and elementary, but heavily uses the specific realization of the $S_n(x)$ through an empirical distribution function, thereby correlating the random variables $S_n(x)$ and $S_n(y)$ properly. This was also crucial for proving Lemma 2. The simple arguments leading to Theorem 1 did not rely on any specific realization.

REFERENCES

1. H. Bauer, Approximation and abstract boundaries, *Amer. Math. Monthly* **85** (1978) 632–647.
2. S. Bernstein, Démonstration du théorème de Weierstrass fondée sur le calcul des probabilities, *Comm. Soc. Math. Kharkov* **13** (1912) 1–2.
3. B. M. Brown, D. Elliott, and D. F. Paget, Lipschitz constants for the Bernstein polynomials of a Lipschitz continuous function, *J. Approx. Theory* **49** (1987) 196–199.
4. P. Diaconis and S. Zabell, Closed form summation for classical distributions: Variations on a theme of De Moivre, *Statist. Science* **6** (1991) 284–302.
5. H. Gzyl and J. L. Palacios, The Weierstrass approximation theorem and large deviations, *Amer. Math. Monthly* **104** (1997) 650–653.
6. M. Kac, Une remarque sur les polynomes de M. S. Bernstein, *Studia Math.* **7** (1938) 49–51.
7. M. Kac, Reconnaissance de priorité relative à ma note, "Une remarque sur les polynomes de M. S. Bernstein", *Studia Math.* **8** (1939) 170.
8. G. G. Lorentz, *Bernstein Polynomials, Math. Expo.* Vol. 8, Univ. of Toronto Press, Toronto, 1953.
9. T. Popoviciu, Sur l'approximation des fonctions convexes d'ordre supérieur, *Mathematica* **10** (1934) 49–54.
10. D. Werner, *Funktionalanalysis, 2nd Ed.* Springer Lehrbuch, Springer, Berlin, 1997.

Weierstrass-Institute for Applied Analysis and Stochastics, Mohrenstrasse 39, D–10117 Berlin, Germany
mathe@wias-berlin.de

An Extension of the Wallace-Simson Theorem: Projecting in Arbitrary Directions

Miguel de Guzmán

1. THE WALLACE-SIMSON LINE. The Wallace line has been a popular object of study for many geometers during the two past centuries. Let us start by recalling the theorem.

The Wallace-Simson Theorem. *Consider a triangle ABC. The locus of all those points P in its plane such that the orthogonal projections of P on the three sides of the triangle are collinear is the circumcircle of ABC. The line of the projections is called the Wallace-Simson line of P with respect to ABC.*

Poncelet attributed this theorem to Robert Simson (1687–1768), among others, since it had the flavor of the geometrical properties that Simson was interested in. Almost certainly it is actually due to Wallace, another Scottish mathematician of lesser importance, who discovered it in 1799.

The beauty of this theorem, which shows a somewhat unexpected and surprising relationship between a triangle and its circumcircle, attracted many geometers in the nineteenth century, among others Jakob Steiner, and led to the discovery of many beautiful properties. Among the most surprising are those connected with Feuerbach's circle, Steiner's deltoid, and Morley's triangle. A reader interested in making an excursion through this landscape can find a guide in the references at the end of this article. Particularly interesting for its numerous references to the older literature is F.G.-M.'s work [**7**, p. 329]. The initials F.G.-M correspond to Fr. Gabriel-Marie, who signed his published works this way.

It is rather simple to obtain a direct proof of the Wallace theorem. One has just to prove that, in Figure 1, one has $\angle BVU = \angle AVW$. Observing that quadrilateral

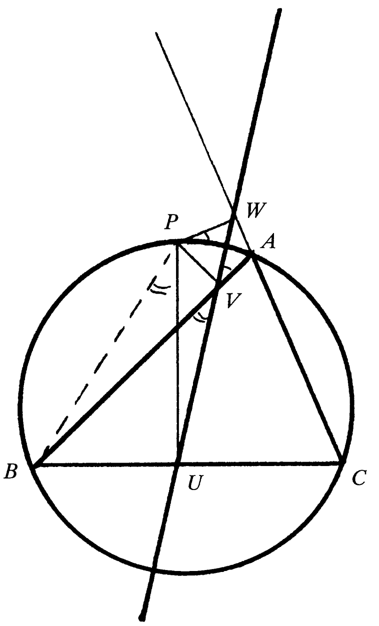


Figure 1

$PVUB$ is cyclic because of the right angles at U and V one gets $\angle BVU = \angle BPU = 90° - \angle PBU = 90° - \angle PBC$. In the same way $PVAW$ is also cyclic and therefore $\angle AVW = \angle APW = 90° - \angle PAW = 90° - (180° - \angle PAC) = \angle PAC - 90°$. But it is quite clear that $\angle PAC - 90° = 90° - \angle PBC$ since the angles at A and B are supplementary in cyclic quadrilateral $PBCA$, and so $\angle BVU = \angle AVW$, which shows that U, V, and W are collinear. ■

2. AN EXTENSION OF THE THEOREM. The goal of this Note is to present a generalization of the Wallace theorem along a strikingly simple direction that I have never seen explored. I first state the theorem and discuss its proof and then present some nice exercises that can be solved in a simple way.

Theorem. *Consider a triangle ABC. Let us call a, b, c, the corresponding sides opposite the vertices. We fix three projection directions α, β, γ, not all three equal, and such that α is not parallel to side a, β is not parallel to side b, and γ is not parallel to side c. Take an arbitrary point P in the plane of ABC and project it on a along α obtaining U, on b along β obtaining V, and on c along γ obtaining W as in Figure 2. Fix a real number k and an orientation in the plane in order to give a sign to the areas of the triangles we consider.*

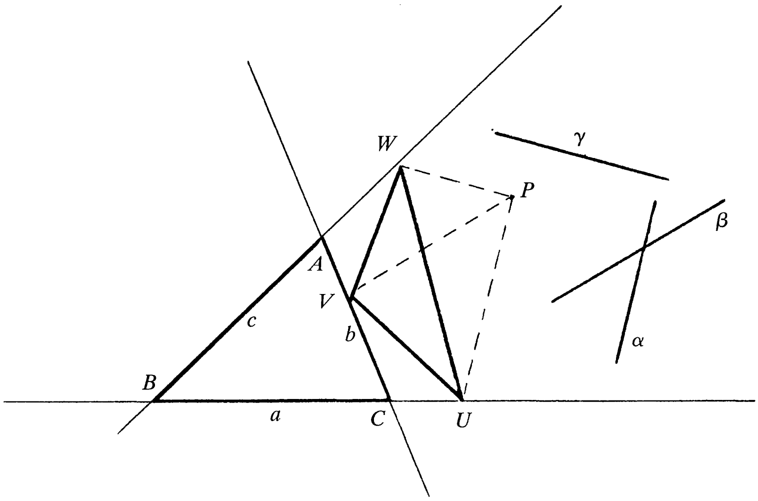


Figure 2

Then the locus of all points P such that the oriented triangle UVW has area k is a conic $\mathbf{C}(k)$. It is clear that $\mathbf{C}(0)$ always goes through the three vertices A, B, C (for example, with P at A, V and W are also at A). The conic $\mathbf{C}(k)$ can, of course, degenerate in various ways.

When k varies (with fixed α, β, γ), the family $\mathbf{C}(k)$ is always a family of conics with the same points at infinity. Furthermore, if one of them has a center, all others have the same center and are homothetical to each other (except the possibly degenerate elements of the family $\mathbf{C}(k)$), the homothecy center being the common center of all such conics. If none has a center, then they are all translations of the same parabola along the direction of its axis.

The construction, with straightedge and compass, of the common center (when it exists) and of the axes and asymptotes of the conics of the family $\mathbf{C}(k)$ is easily done once one knows A, B, C and the projection directions α, β, γ.

Proof: The theorem is one of those results whose only difficulty is arriving at its statement, since the easy analytical proof we develop could be left as an exercise to the reader.

Let us begin by fixing an arbitrary cartesian system of coordinates and an orientation to give a sign to the areas of the triangles. If the point P has coordinates (x, y) and we denote by α, β, γ three vectors that correspond to the given projection directions, it is clear that the points U, V, W have as coordinates, respectively,

$$(u_1(x,y), u_2(x,y)), (v_1(x,y), v_2(x,y)), (w_1(x,y), w_2(x,y))$$

where each of these functions is a linear function in x, y with coefficients that depend only on the parameters (already fixed) that determine $A, B, C, \alpha, \beta, \gamma$, in our coordinate system.

The area of the triangle UVW is given by half the value of the determinant of the matrix

$$\begin{pmatrix} u_1(x,y) & u_2(x,y) & 1 \\ v_1(x,y) & v_2(x,y) & 1 \\ w_1(x,y) & w_2(x,y) & 1 \end{pmatrix}.$$

Therefore the equation of the locus is of the form

$$mx^2 + ny^2 + 2pxy + 2qx + 2ry + s = 2k \tag{1}$$

where m, n, p, q, r, s depend only on the fixed entities $A, B, C, \alpha, \beta, \gamma$. This shows that $\mathbf{C}(k)$ is a conic and that, when k varies, $\mathbf{C}(k)$ is a family of conics whose points at infinity are the same, since they are determined by

$$mx^2 + ny^2 + 2pxy = 0. \tag{2}$$

The equation (2) cannot degenerate, i.e., m, n, p cannot be 0 simultaneously. In fact, in that case $\mathbf{C}(0)$ would have as equation $2qx + 2ry + s = 0$. If one of the coefficients q, r is not zero, then $\mathbf{C}(0)$ is a straight line, which is false, since A, B, C are points of $\mathbf{C}(0)$. If $q = r = 0$ then the equation of $\mathbf{C}(0)$ is $s = 0$. Therefore, if s is not 0, $\mathbf{C}(0)$ is empty, which is again false. If $s = 0$ then $\mathbf{C}(0) = \mathbb{R}^2$, which easily implies that all three directions α, β, γ, are the same; this was excluded from the beginning.

Therefore, $\mathbf{C}(k)$ is a family of conics with the same points at infinity, as announced in the statement of the theorem. If we take as the origin an arbitrary point O and as coordinate axes the two (always real) bisectors of the angles formed by the lines joining O with the points at infinity, then the equation (1) takes the form

$$mx^2 + ny^2 + 2qx + 2ry + s = 2k \tag{3}$$

(where m, n, q, r, s, of course, need not be the same as before).

If for some k the conic $\mathbf{C}(k)$ has a point Z as center and we take Z as the origin and as coordinate axes the axes of the conic (the two bisectors of the angle obtained by joining Z to the two points at infinity), then the equation (3) takes the form

$$mx^2 + ny^2 + s = 2k \tag{4}$$

with m, n, s depending only, as before, on $A, B, C, \alpha, \beta, \gamma$, which makes plain the simple structure of the family $\mathbf{C}(k)$. All, except for $s = 2k$, are concentric and homothetic with respect to Z. For $s = 2k$ the conic $\mathbf{C}(k)$ degenerates and becomes either a pair of real lines (which can coincide) or a pair of imaginary lines that intersect at Z.

If in equation (4) it happens, for example, that $m = 0$ then the equation of $\mathbf{C}(k)$ would be $ny^2 + s = 2k$, which is, for each k, a pair of parallel lines, real or complex according to the value of k; this pair of lines becomes one double real line if $s = 2k$. It is clear, in this case, that any point of such a double line is a center of symmetry of each $\mathbf{C}(k)$.

If no $\mathbf{C}(k)$ has a center, then in the equation

$$mx^2 + ny^2 + 2qx + 2ry + s = 2k$$

we have either $m = 0$ or $n = 0$. Assume $n = 0$. Then

$$mx^2 + 2qx + 2ry + s = 2k$$

and if we make the change $x = X - q/m$ (i.e., a translation of the y axis) we obtain

$$mX^2 + 2ry + t = 2k$$

with m, r, t depending only, as before, on $A, B, C, \alpha, \beta, \gamma$. The coefficient r cannot be null, since if $r = 0$ we would have $mX^2 + t = 2k$, and this is, for each k, a pair of parallel lines and thus has a center. Therefore, r is not null and so $\mathbf{C}(k)$ is, in this case, a family of parabolas which are all obtained as translations of one of them in the direction of its axis.

Construction, with compass and straightedge, of the center, axes and asymptotes of $\mathbf{C}(k)$.

Here we indicate a practical construction for the main elements of the conics $\mathbf{C}(k)$. According to the previous part of the theorem it is sufficient to find these elements for $\mathbf{C}(0)$, of which we already know three points A, B, C.

If two of the projection directions coincide, for example $\beta = \gamma$, it is easy to show in a direct way that $\mathbf{C}(0)$ is the line BC together with the line through A in the direction $\beta = \gamma$, independent of the direction α. It is also easy to show that coincidence of two of the projection directions is a necessary and sufficient condition for degeneration of the conic $\mathbf{C}(0)$ into two lines (one of them is the line containing one of the sides of the triangle and the other is the line through the opposite vertex in the same direction as the two coinciding ones). It is then quite clear that, if $\beta = \gamma$ does not coincide with the direction of the side BC, then $\mathbf{C}(k)$, with k different from 0, is a family of hyperbolas having these two lines as asymptotes that are homothetic to each other with respect to their common center, the intersection of those two lines. When $\beta = \gamma$ has the same direction as the side BC, each $\mathbf{C}(k)$ is a pair of parallel lines (that coincide for $k = 1/2$) and parallel to those that constitute $\mathbf{C}(0)$. Thus, we have determined the main elements of $\mathbf{C}(k)$ if $\mathbf{C}(0)$ degenerates.

If no pair of projection directions coincide, then $\mathbf{C}(0)$ is a non-degenerate conic for which we already know three points. Another three are easily determined in the following way:

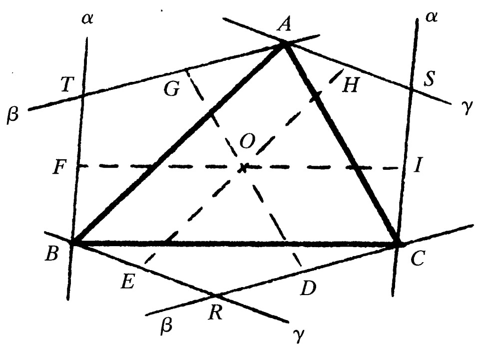


Figure 3

Through B we draw a line parallel to γ and through C a line parallel to β; this makes $V = C$ and $W = B$, and UVW degenerates; it is clear that the intersection R belongs to $\mathbf{C}(0)$. Through C we draw a line parallel to α and through A a line parallel to γ; similarly the intersection S belongs to $\mathbf{C}(0)$. Through A we draw a line parallel to β and through B a line parallel to α; the intersection T belongs to $\mathbf{C}(0)$.

We thus obtain a hexagon $ATBRCS$ whose vertices are in $\mathbf{C}(0)$ and such that its opposite sides are parallel. It is clear that the midpoints of the sides of a pair of opposite sides are on a diameter of $\mathbf{C}(0)$. We thus obtain the center of $\mathbf{C}(0)$ (when it does exist) as intersection of these three diameters, DG, EH, FI (see Figure 3). When the center does not exist we obtain at least the direction of the axis of the parabola $\mathbf{C}(0)$.

Assume for the moment that there is a center as indicated in Figure 3. Since, we already have three pairs of conjugate diameters, the construction, with compass and straightedge, of the common asymptotes and axes of all conics $\mathbf{C}(k)$ is a well known exercise.

For the determination of the vertex of the parabola $\mathbf{C}(0)$ when the center does not exist, we proceed as follows. Through A we draw a line perpendicular to the direction of the axis, which we already know. We determine its intersection A' with the parabola; the bisector line of AA' is the axis of $\mathbf{C}(0)$. The intersection of this line with the parabola gives us the vertex. The parabolas $\mathbf{C}(k)$ are the translations of $\mathbf{C}(0)$ in the direction of its axis.

3. SOME INTERESTING EXERCISES RELATED TO THE THEOREM. The theorem just proved suggests many interesting exercises, which can be solved easily with the methods we have used in the proof. Here are a few with some indications of the way to solve them. Fix a triangle ABC.

1. *Find necessary and sufficient conditions on the projection directions in order that $\mathbf{C}(0)$ and $\mathbf{C}(k)$ are circles. (Of course, $\mathbf{C}(0)$ will be the circumcircle and all $\mathbf{C}(k)$ will be concentric to it.)*

Hint: Consider the inscribed hexagon in $\mathbf{C}(0)$ that we used in the construction of the main elements of $\mathbf{C}(0)$. The orthogonal projections to the sides (the Wallace Theorem) are not the only possibilities in order to get circles.

2. *Determine necessary and sufficient conditions on the projection directions in order that the conics $\mathbf{C}(k)$ are equilateral hyperbolas.*

Hint: Remember that a necessary and sufficient condition for a conic that goes through three points A, B, C, to be an equilateral hyberbola is that it also goes through the orthocenter of ABC.

3. *For two different projection directions, determine a third projection so that the $\mathbf{C}(k)$ are equilateral hyperbolas.*

Hint: See the hint for Exercise 2.

4. *Given three directions α, β, γ, find, if possible, a point F such that the triangle MNP with vertices at the projections of F on the sides a, b, c in the directions α, β, γ has maximum or minimum area.*

Hint: Consider the equation of the conic $\mathbf{C}(k)$ with respect to its axes.

5. *Given a non-degenerate conic $\mathbf{C}$ and the triangle ABC inscribed in it, determine three directions α, β, γ such that $\mathbf{C}$ is the conic $\mathbf{C}(0)$ for ABC and the three given directions.*

Hint: Remember the hexagon inscribed in the conic that we used in the construction of the elements of $\mathbf{C}(0)$.

6. *Assume that triangle ABC has area S and that the radius of its circumscribed circle G is R. We draw a circle K concentric with G and with radius r. From a point P of K we draw its projections U, V, W on the sides of ABC. Determine, as a function of S, R, and r, the area of the triangle UVW.*

Hint: The same as in Exercise 4. Answer: Area(UVW) $= (S/4)(1 - r^2/R^2)$, *having selected the appropriate orientation so that the triangle UVW has positive area when* $r < R$.

REFERENCES

1. H. F. Baker, *An Introduction to Plane Geometry*, Cambridge University Press, London, 1963.
2. J. H. Butchart, The Deltoid Regarded as the Envelope of Simson Lines, *Amer. Math. Monthly* **46** (1939) 85–86
3. Shang-Ching Chou, Proving Elementary Geometry Theorems Using Wu's Algorithm, *Contemporary Mathematics* **29** (1984) 243–286.
4. H. S. M. Coxeter, *Introduction to Geometry*, John Wiley, New York, 1969.
5. H. S. M. Coxeter and S. L. Greitzer, *Geometry Revisited*, Mathematical Association of America, 1967.
6. H. Dörrie, *100 Great Problems of Elementary Mathematics: Their History and Solution*, Dover, New York, 1965.
7. F. G.-M., *Exercises de Géométrie*, Maison, Tours, 1912.
8. B. C. Patterson, The Triangle: its Deltoids and Foliates, *Amer. Math. Monthly* **47** (1940) 11–18.
9. O. J. Ramler, The Orthopole Loci of Some One-Parameter Systems of Lines Referred to a Fixed Triangle, *Amer. Math. Monthly* **37** (1930) 130–136.
10. C. E. Van Horn, The Simson Quartic of a Triangle, *Amer. Math. Monthly* **45** (1938) 434–437.
11. D. Wells, *The Penguin Dictionary of Curious and Interesting Geometry*, Penguin, London, 1991.

Facultad de Matemáticas, Universidad Complutense de Madrid, 28040 Madrid, Spain
mdeguzman@bitmailer.net

Another Short Proof of Ramanujan's Mod 5 Partition Congruence, and More

Michael D. Hirschhorn

We present another novel short proof of Ramanujan's partition congruence

$$p(5n + 4) \equiv 0 \pmod 5 \tag{1}$$

in addition to that presented by John L. Drost [**2**], and indeed prove rather more.

Ramanujan made the remarkable observation from a table of values of $p(n)$, the number of partitions of n, that $p(5n + 4)$ is divisible by 5. He observed and conjectured much more, and his conjectures turned out in the main to be correct. He gave a simple proof, based upon identities of Euler and Jacobi, of the conjecture (1), and his proof is essentially the one reproduced in Hardy and Wright [**3**] and referred to by Drost. Ramanujan's proof relies on manipulating power series, and considering coefficients modulo 5. It is my intention to give a proof of a

similar sort, more transparent than that of Ramanujan, using only the identity of Jacobi. And further, with a little extra work including the use of Jacobi's triple-product identity, we prove remarkable congruences for the partition function due to Atkin and Swinnerton-Dyer.

As is usual, write $(q)_\infty = \prod_{n\geq 1}(1-q^n)$. Then

$$\sum_{n\geq 0} p(n)q^n = \frac{1}{(q)_\infty}.$$

We begin with Jacobi's identity [**3**, Theorem 357],

$$(q)_\infty^3 = \sum_{n\geq 0} (-1)^n(2n+1)q^{n(n+1)/2}.$$

Each coefficient is congruent modulo 5 to 0, ± 1 or ± 2. Specifically, the coefficient is congruent to 1 when $n \equiv 0$ or 9 (mod 10), -1 when $n \equiv 4$ or 5 (mod 10), $+2$ when $n \equiv 1$ or 8 (mod 10), -2 when $n \equiv 3$ or 6 (mod 10), and 0 when $n \equiv 2$ or 7 (mod 10). Thus we find that, modulo 5,

$$\begin{aligned}
(q)_\infty^3 \equiv{}& \sum_{n\geq 0} q^{10n(10n+1)/2} - \sum_{n\geq 0} q^{(10n+4)(10n+5)/2} - \sum_{n\geq 0} q^{(10n+5)(10n+6)/2}\\
&+ \sum_{n\geq 0} q^{(10n+9)(10n+10)/2} + 2\sum_{n\geq 0} q^{(10n+1)(10n+2)/2} - 2\sum_{n\geq 0} q^{(10n+3)(10n+4)/2}\\
&- 2\sum_{n\geq 0} q^{(10n+6)(10n+7)/2} + 2\sum_{n\geq 0} q^{(10n+8)(10n+9)/2}\\
\equiv{}& \sum_{n\geq 0} q^{50n^2+5n} - \sum_{n\geq 0} q^{50n^2+45n+10} - \sum_{n\geq 0} q^{50n^2+55n+15} + \sum_{n\geq 0} q^{50n^2+95n+45}\\
&+ 2\sum_{n\geq 0} q^{50n^2+15n+1} - 2\sum_{n\geq 0} q^{50n^2+35n+6} - 2\sum_{n\geq 0} q^{50n^2+65n+21}\\
&+ 2\sum_{n\geq 0} q^{50n^2+85n+36}.
\end{aligned}$$

Observe that in the first four sums the powers of q are congruent to 0 (mod 5) while in the latter four sums the powers of q are congruent to 1 (mod 5). Thus we have

$$(q)_\infty^3 \equiv X + 2qY,$$

where each of X, Y is a series in powers of q^5.

Also

$$\begin{aligned}
(q)_\infty^5 &= \prod_{n\geq 1}(1-q^n)^5 = \prod_{n\geq 1}(1 - 5q^n + 10q^{2n} - 10q^{3n} + 5q^{4n} - q^{5n})\\
&\equiv \prod_{n\geq 1}(1-q^{5n}) \equiv (q^5)_\infty.
\end{aligned}$$

Thus

$$\begin{aligned}
\sum_{n\geq 0} p(n)q^n &= \frac{1}{(q)_\infty} = \frac{(q)_\infty^9}{(q)_\infty^{10}} = \frac{\left((q)_\infty^3\right)^3}{\left((q)_\infty^5\right)^2} \equiv \frac{\left((q)_\infty^3\right)^3}{(q^5)_\infty^2} \equiv \frac{(X+2qY)^3}{(q^5)_\infty^2}\\
&\equiv \frac{X^3 + 6qX^2Y + 12q^2XY^2 + 8q^3Y^3}{(q^5)_\infty^2}\\
&\equiv \frac{X^3 + qX^2Y + 2q^2XY^2 + 3q^3Y^3}{(q^5)_\infty^2}.
\end{aligned}$$

Comparing terms containing powers of q congruent to 4 modulo 5 on both sides, we see that

$$\sum_{n \geq 0} p(5n+4) q^{5n+4} \equiv 0 \pmod{5}. \qquad \blacksquare$$

Notice that, at no extra cost, we obtain the congruences

$$\sum_{n \geq 0} p(5n) q^{5n} \equiv X^3 / (q^5)_\infty^2,$$

$$\sum_{n \geq 0} p(5n+1) q^{5n+1} \equiv q X^2 Y / (q^5)_\infty^2,$$

$$\sum_{n \geq 0} p(5n+2) q^{5n+2} \equiv 2 q^2 X Y^2 / (q^5)_\infty^2, \quad \text{and}$$

$$\sum_{n \geq 0} p(5n+3) q^{5n+3} \equiv 3 q^3 Y^3 / (q^5)_\infty^2.$$

It is not hard to show that each of X, Y is an infinite product. Indeed, as we shall see,

$$X = \prod_{n \geq 1} (1 - q^{25n-15})(1 - q^{25n-10})(1 - q^{25n}), \tag{2}$$

$$Y = \prod_{n \geq 1} (1 - q^{25n-20})(1 - q^{25n-5})(1 - q^{25n}). \tag{3}$$

It follows that

$$\sum_{n \geq 0} p(5n) q^n \equiv \prod_{n \geq 1} \frac{(1 - q^{5n-3})(1 - q^{5n-2})(1 - q^{5n})}{(1 - q^{5n-4})^2 (1 - q^{5n-1})^2},$$

$$\sum_{n \geq 0} p(5n+1) q^n \equiv \prod_{n \geq 1} \frac{(1 - q^{5n})}{(1 - q^{5n-4})(1 - q^{5n-1})},$$

$$\sum_{n \geq 0} p(5n+2) q^n \equiv 2 \prod_{n \geq 1} \frac{(1 - q^{5n})}{(1 - q^{5n-3})(1 - q^{5n-2})}, \quad \text{and}$$

$$\sum_{n \geq 0} p(5n+3) q^n \equiv 3 \prod_{n \geq 1} \frac{(1 - q^{5n-4})(1 - q^{5n-1})(1 - q^{5n})}{(1 - q^{5n-3})^2 (1 - q^{5n-2})^2}.$$

These remarkable results are due to Atkin and Swinnerton-Dyer [**1**, Theorem 1].

We now show that X, Y are the infinite products claimed in (2) and (3). We have

$$X = \sum_{n \geq 0} q^{50n^2+5n} - \sum_{n \geq 0} q^{50n^2+45n+10} - \sum_{n \geq 0} q^{50n^2+55n+15} + \sum_{n \geq 0} q^{50n^2+95n+45}$$

In the first sum, replace n by $-n$, in the third replace n by $-n-1$, and in the fourth replace n by $n-1$. Then we find

$$\begin{aligned} X &= \sum_{n \leq 0} q^{50n^2-5n} - \sum_{n \geq 0} q^{50n^2+45n+10} - \sum_{n \leq -1} q^{50n^2+45n+10} + \sum_{n \geq 1} q^{50n^2-5n} \\ &= \sum_{n=-\infty}^{\infty} q^{50n^2-5n} - \sum_{n=-\infty}^{\infty} q^{50n^2+45n+10} \\ &= \sum_{n=-\infty}^{\infty} (-1)^n q^{(25n^2-5n)/2}. \end{aligned}$$

The terms for n even in the final sum correspond to the first sum on the line above; the terms for n odd to the second sum.

In the same way, we find

$$\begin{aligned} Y &= \sum_{n\geq 0} q^{50n^2+15n} - \sum_{n\geq 0} q^{50n^2+35n+5} - \sum_{n\geq 0} q^{50n^2+65n+20} + \sum_{n\geq 0} q^{50n^2+85n+35} \\ &= \sum_{n\leq 0} q^{50n^2-15n} - \sum_{n\geq 0} q^{50n^2+35n+5} - \sum_{n\leq -1} q^{50n^2+35n+5} + \sum_{n\geq 1} q^{50n^2-15n} \\ &= \sum_{n=-\infty}^{\infty} q^{50n^2-15n} - \sum_{n=-\infty}^{\infty} q^{50n^2+35n+5} \\ &= \sum_{n=-\infty}^{\infty} (-1)^n q^{(25n^2-15n)/2}. \end{aligned}$$

To complete the proof, we now invoke Jacobi's triple product identity [**3**, Theorem 352], in the form

$$\sum_{n=-\infty}^{\infty} (-1)^n a^n q^{(n^2-n)/2} = \prod_{n\geq 1} (1 - aq^{n-1})(1 - a^{-1}q^n)(1 - q^n)$$

with q replaced by q^{25} and a replaced by q^{10} and q^5, respectively.

REFERENCES

1. A. O. L. Atkin and P. Swinnerton-Dyer, Some properties of partitions, *Proc. London Math. Soc.* **4** (1954), 84–106.
2. John L. Drost, A shorter proof of the Ramanujan congruence mod 5, this MONTHLY **104** (1997) 963–964.
3. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 5th edition, Clarendon Press, Oxford 1979.

School of Mathematics, UNSW, Sydney 2052, NSW, Australia
m.hirschhorn@unsw.edu.au

UNSOLVED PROBLEMS

Edited by **Richard Nowakowski**

In this department the MONTHLY presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial or related results. Typescripts should be sent to Richard Nowakowski, Department of Mathematics & Statistics & Computing Science, Dalhousie University, Halifax NS, Canada B3H 3J5, rjn@mscs.dal.ca

The Cayley Addition Table of Z_n

Hunter S. Snevily

Few mathematical objects could be considered more simple than the Cayley addition table of $\mathbf{Z}_n$ but we show that even these simple objects have some interesting yet unproved properties.

A *transversal* of an $n \times n$ matrix is a collection of n cells, no two of which are in the same row or column. A transversal of a matrix is a *latin* transversal if no two of its cells contain the same element.

Conjecture 1. For any odd n and any $k \in \{1, \dots, n\}$, any $k \times k$ submatrix of the Cayley addition table of $\mathbf{Z}_n$ contains a latin transversal.

Conjecture 2. For any even n and any $k \in \{1, \dots, n\}$, and any $k \times k$ submatrix of the Cayley addition table of $\mathbf{Z}_n$ contains a latin transversal provided the submatrix is not a subgroup of even order or a translate of such a subgroup.

Perhaps a stronger version of Conjecture 1 might be true.

Conjecture 3. Let A be the Cayley table of any Abelian group of odd order n, and let $k \in \{1, \dots, n\}$. Then any $k \times k$ submatrix of A contains a latin transversal.

Conjecture 3 implies a fundamental property of finite fields of odd characteristic: Let $\mathbf{GF}[\mathbf{q}]$ be a finite field of odd characteristic containing q elements and let $A = \{a_1, a_2, \dots, a_k\}$ and $B = \{b_1, b_2, \dots, b_k\}$, $k \leq q$, be two subsets of $\mathbf{GF}[\mathbf{q}]$. Then there is a permutation $\pi \in S_k$ such that the sums $a_i + b_{\pi(i)}$, in ($\mathbf{GF}[\mathbf{q}]$), are pairwise distinct.

So far the only result that supports these conjectures (and the next) is the following theorem of Noga Alon. The special case $k = n$ for Conjectures 1, 2, and 3 was proved in **[2]**.

Theorem 1 (Alon). *Let p be a prime, suppose $k < p$, let $(a_1, a_2, \dots, a_k)$ be a sequence of not necessarily distinct members of the finite field $\mathbf{Z}_p$, and let B be a subset*

of cardinality k of $\mathbf{Z}_p$. Then there is a numbering $(b_1, b_2, \dots, b_k)$ of the elements of B such that the sums $a_i + b_i$ (in $\mathbf{Z}_p$) are pairwise distinct.

Consider the infinitely long periodic sequence of integers modulo n. For example, modulo 6 the sequence is: 012345012345012345.... Now take k copies of this sequence ($k < n$) and start them at any position (i.e., truncate them anywhere).

For example, taking the sequence modulo 6 and $k = 3$,

$$\begin{array}{l} 012345012345012345\dots; \\ 012345012345012345\dots; \\ 012345012345012345\dots \end{array}$$

all start at the same position and

$$\begin{array}{l} 012345012345012345\dots; \\ 234501234501234501\dots; \\ 501234501234501234\dots. \end{array}$$

all start at different positions. A $k \times k$ *frame* (matrix) consists of k consecutive columns taken from such an arrangement. For the second example all frames look like:

$$\begin{array}{c} abc \\ cef \\ gab \end{array}$$

This is the *template* of our k sequences.

Conjecture 4. Take k copies of the infinitely long periodic sequence modulo n (k < n) and start them anywhere (i.e., truncate them anywhere). Then the template has a latin transversal.

Note that Alon has proved this conjecture when n is a prime. The following examples show why k cannot equal n:

$n = k = 4$		$n = k = 5$
0123		01234
0123	and	01234
0123		01234
1230		01234
		12340

REFERENCES

1. N. Alon, *Additive Latin Transversals*, preprint.
2. M. Hall, Jr., A combinatorial problem on abelian groups, *Proc. Amer. Math. Soc.* **3** (1952) 584–587.

University of Idaho, Moscow, ID 83844-1103
snevily@uidaho.edu

PROBLEMS AND SOLUTIONS

Edited by **Gerald A. Edgar, Daniel H. Ullman,** and **Douglas B. West**
with the collaboration of Paul T. Bateman, Mario Benedicty, Paul Bracken, Duane M. Broline, Ezra A. Brown, Richard T. Bumby, Glenn G. Chappell, Randall Dougherty, Roger B. Eggleton, Ira M. Gessel, Bart Goddard, Jerrold R. Griggs, Douglas A. Hensley, Richard Holzsager, John R. Isbell, Robert Israel, Kiran S. Kedlaya, Murray S. Klamkin, Fred Kochman, Frederick W. Luttmann, Vania Mascioni, Frank B. Miles, Richard Pfiefer, Cecil C. Rousseau, Leonard Smiley, John Henry Steelman, Kenneth Stolarsky, Richard Stong, Charles Vanden Eynden, and William E. Watkins.

Proposed problems and solutions should be sent in duplicate to the MONTHLY *problems address on the inside front cover. Submitted problems should include solutions and relevant references. Submitted solutions should arrive at that address before November 30, 1999; Additional information, such as generalizations and references, is welcome. The problem number and the solver's name and address should appear on each solution. An acknowledgement will be sent only if a mailing label is provided. An asterisk (*) after the number of a problem or a part of a problem indicates that no solution is currently available.*

PROBLEMS

10739. *Proposed by Oscar Ciaurri, Logroño, Spain.* Suppose that $f\colon [0,1] \to \mathbb{R}$ has a continuous second derivative with $f''(x) > 0$ on $(0,1)$, and suppose that $f(0) = 0$. Choose $a \in (0,1)$ such that $f'(a) < f(1)$. Show that there is a unique $b \in (a,1)$ such that $f'(a) = f(b)/b$.

10740. *Proposed by Charles Vanden Eynden, Illinois State University, Normal, IL.* A connected bipartite simple graph has a unique bipartition, meaning a partition of the vertices into two independent sets. Let **G** be the set of such graphs that have no isomorphism that interchanges the two sets of the bipartition. Is there a criterion that for each $G \in \mathbf{G}$ selects a well-defined set of the bipartition?

10741. *Proposed by Tim Keller, Fair Oaks, CA.* Is there an even base b for which there exist square integers of the form $dddd_b$? By $dddd_b$, we mean the four-digit number in base b all of whose digits are d. For odd b we have the examples $1111_7 = 20^2$ and $4444_7 = 40^2$.

10742. *Proposed by Emre Alkan, University of Wisconsin, Madison, WI.* Let us say that a finite group G has the *maximal property* if, for any prime p that divides $|G|$, G has a maximal subgroup H such that $p|H|$ divides $|G|$.
(a) Show that every finite solvable group has the maximal property.
(b) Show that there are infinitely many finite groups with the maximal property that are not solvable.
(c) Show that there are infinitely many finite groups without the maximal property that are not solvable.

10743. *Proposed by Călin Popescu, Université Catholique de Louvain, Louvain-La-Neuve, Belgium.* Let $p \geq 5$ be prime, and let n be an integer such that $(p+1)/2 \leq n \leq p-2$. Let $R = \sum(-1)^i \binom{n}{i}$, where the sum is taken over the quadratic residues i modulo p, and let $N = \sum(-1)^j \binom{n}{j}$, where the sum is taken over the quadratic nonresidues j modulo p. Prove that exactly one of R and N is divisible by p.

10744. *Proposed by Peter Lindqvist, Norwegian University of Science and Technology, Trondheim, Norway, and Jaak Peetre, University of Lund, Lund, Sweden.* Fix $p > 0$, and define functions $S(x)$, $C(x)$, and $T(x)$ for sufficiently small x by

$$x = \int_0^{S(x)} \frac{dt}{(1-t^p)^{(p-1)/p}}, \quad x = \int_{C(x)}^{1} \frac{dt}{(1-t^p)^{(p-1)/p}}, \quad x = \int_0^{T(x)} \frac{dt}{(1+t^p)^{2/p}}.$$

Show that $S(x)^p + C(x)^p = 1$ and that $T(x) = S(x)/C(x)$. The case $p = 2$ yields the familiar trigonometric formulas.

10745. *Proposed by M. J. Pelling, London, England.* For $n \geq 1$, let $f(n)$ be the number of solutions (r, s, t) in positive integers to the Diophantine equation $rst = n(r+s+t)$.
(a) Prove that $f(n) = O(n^{1/2+\delta})$ for every $\delta > 0$.
(b)* Prove that $f(n) = O(n^{\delta})$ for every $\delta > 0$.

SOLUTIONS

Using the Walls to Find the Center

10386 [1994, 474]. *Proposed by Jordan Tabov, Bulgarian Academy of Sciences, Sofia, Bulgaria.* Let a tetrahedron with vertices A_1, A_2, A_3, A_4 have altitudes that meet in a point H. For any point P, let P_1, P_2, P_3, and P_4 be the feet of the perpendiculars from P to the faces $A_2A_3A_4$, $A_3A_4A_1$, $A_4A_1A_2$, and $A_1A_2A_3$, respectively. Prove that there exist constants a_1, a_2, a_3, and a_4 such that one has

$$a_1\overrightarrow{PP_1} + a_2\overrightarrow{PP_2} + a_3\overrightarrow{PP_3} + a_4\overrightarrow{PP_4} = \overrightarrow{PH}$$

for every point P.

Solution by Murray S. Klamkin, University of Alberta, Edmonton, Alberta, Canada. More generally, let H and P be any two points in the space of the given tetrahedron and let P_1, P_2, P_3, P_4 be the feet of the lines through P parallel to HA_1, HA_2, HA_3, HA_4 in the faces of the tetrahedron opposite A_1, A_2, A_3, A_4, respectively. Then there exist constants a_1, a_2, a_3, a_4, independent of P, such that

$$a_1\overrightarrow{PP_1} + a_2\overrightarrow{PP_2} + a_3\overrightarrow{PP_3} + a_4\overrightarrow{PP_4} = \overrightarrow{PH}.$$

Let $\mathbf{V}$ denote the vector from an origin outside the space of the given tetrahedron to any point V in the space of the tetrahedron. Then H and P have the representations (barycentric coordinates)

$$\mathbf{H} = x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + x_3\mathbf{A}_3 + x_4\mathbf{A}_4 \quad (x_1+x_2+x_3+x_4 = 1),$$

$$\mathbf{P} = u_1\mathbf{A}_1 + u_2\mathbf{A}_2 + u_3\mathbf{A}_3 + u_4\mathbf{A}_4 \quad (u_1+u_2+u_3+u_4 = 1).$$

Since P_1 has the representation $\mathbf{P}_1 = r_2\mathbf{A}_2 + r_3\mathbf{A}_3 + r_4\mathbf{A}_4$, where $r_2 + r_3 + r_4 = 1$, we must have

$$r_2\mathbf{A}_2 + r_3\mathbf{A}_3 + r_4\mathbf{A}_4 - \mathbf{P} = \lambda_1(\mathbf{H} - \mathbf{A}_1).$$

Since $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$, $\mathbf{A}_4$ are independent vectors, we get $\lambda_1 = u_1/(1-x_1)$, so that $\overrightarrow{PP_1} = (\mathbf{P}_1 - \mathbf{P}) = (\mathbf{H} - \mathbf{A}_1)u_1/(1-x_1)$. Similarly,

$$(\mathbf{P}_i - \mathbf{P}) = (\mathbf{H} - \mathbf{A}_i)\frac{u_i}{1-x_i} \quad \text{for } i = 1, 2, 3, 4.$$

Choosing $a_i = 1 - x_i$, we obtain

$$\sum a_i(\mathbf{P}_i - \mathbf{P}) = \sum u_i(\mathbf{H} - \mathbf{A}_i) = \mathbf{H} - \mathbf{P} = \overrightarrow{PH}.$$

This proof generalizes to give an analogous result for n-dimensional simplices.

Solved also by J. Anglesio (France), R. J. Chapman (U. K.), M. Golomb, K. Hanes, N. Komanda, O. P. Lossers (The Netherlands), and the proposer.

A Reciprocal Summation Identity

10490 [1995, 930]. *Proposed by Seung-Jin Bang, Ajou University, Suwon, Korea.* Show that

$$\sum_{k=1}^{n}\frac{(-1)^{k-1}}{k}\binom{n}{k}\sum_{j=1}^{k}\frac{1}{j}\left(\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{j}\right)=\sum_{k=1}^{n}\frac{1}{k^3}$$

for each positive integer n.

Solution I by Jeremy E. Dawson, Australian National University, Canberra, Australia. Since

$$\int_0^1\frac{1}{z}\int_0^z\frac{1}{y}\int_0^y x^{k-1}\,dx\,dy\,dz=\frac{1}{k^3},$$

we have $\sum_{k=1}^{n}1/k^3=\int_0^1 z^{-1}\int_0^z y^{-1}\int_0^y(1-x^n)/(1-x)\,dx\,dy\,dz$. Letting $w=1-x$, $v=1-y$, and $u=1-z$, we can rewrite the latter integral as

$$\int_0^1\frac{1}{1-u}\int_u^1\frac{1}{1-v}\int_v^1\frac{1-(1-w)^n}{w}dw\,dv\,du.$$

Now use $\left(1-(1-w)^n\right)/w=\sum_{k=1}^{n}(-1)^{k-1}\binom{n}{k}w^{k-1}$ and interchange the order of summation and integration. For the resulting multiple integrals, we use

$$\frac{1}{1-s}\int_s^1 t^{m-1}dt=\frac{1-s^m}{m(1-s)}=\frac{1}{m}\sum_{l=1}^{m}s^{l-1}$$

twice to obtain

$$\int_0^1\frac{1}{1-u}\int_u^1\frac{1}{1-v}\int_v^1 w^{k-1}dw\,dv\,du=\int_0^1\frac{1}{k}\sum_{j=1}^{k}\left(\frac{1}{j}\sum_{i=1}^{j}u^{i-1}\right)du=\frac{1}{k}\sum_{j=1}^{k}\frac{1}{j}\sum_{i=1}^{j}\frac{1}{i}.$$

Thus

$$\sum_{k=1}^{n}\frac{1}{k^3}=\sum_{k=1}^{n}(-1)^{k-1}\binom{n}{k}\frac{1}{k}\sum_{j=1}^{k}\frac{1}{j}\sum_{i=1}^{j}\frac{1}{i}.$$

Solution II by A. N. 't Woord, University of Technology, Eindhoven, The Netherlands. We prove first that if $b_n=\sum_{k=1}^{n}a_k\binom{n}{k}$ for $n\geq 1$, then

$$(1)\quad\sum_{k=1}^{n}\frac{b_k}{k}=\sum_{k=1}^{n}\frac{a_k}{k}\binom{n}{k},\qquad\text{and}\qquad(2)\quad a_n=\sum_{k=1}^{n}(-1)^{n-k}\binom{n}{k}b_k\ \text{ for } n\geq 1.$$

(1) follows by induction on n: it is immediate for $n=1$; and for $n>1$ we have

$$\sum_{k=1}^{n}\frac{a_k}{k}\binom{n}{k}=\sum_{k=1}^{n}\frac{a_k}{k}\binom{n-1}{k}+\sum_{k=1}^{n}\frac{a_k}{k}\binom{n-1}{k-1}=\sum_{k=1}^{n-1}\frac{b_k}{k}+\sum_{k=1}^{n}\frac{a_k}{n}\binom{n}{k}=\sum_{k=1}^{n}\frac{b_k}{k}.$$

For (2), we use $\binom{n}{k}\binom{k}{i}=\binom{n}{i}\binom{n-i}{n-k}$ and $\sum_{k=i}^{n}(-1)^{n-k}\binom{n-i}{n-k}=\delta_{in}$ to obtain

$$\sum_{k=1}^{n}(-1)^{n-k}\binom{n}{k}b_k=\sum_{k=1}^{n}\sum_{i=1}^{k}(-1)^{n-k}\binom{n}{k}\binom{k}{i}a_i=\sum_{i=1}^{n}a_i\binom{n}{i}\sum_{k=i}^{n}(-1)^{n-k}\binom{n-i}{n-k}=a_n.$$

We now begin with the identity $1=\sum_{k=1}^{n}(-1)^{k-1}\binom{n}{k}$. Applying (1) twice yields

$$\sum_{k=1}^{n}\frac{1}{k}\sum_{j=1}^{k}\frac{1}{j}=\sum_{k=1}^{n}\frac{(-1)^{k-1}}{k^2}\binom{n}{k}.$$

Applying (2) now yields

$$\frac{(-1)^{n-1}}{n^2} = \sum_{k=1}^{n}(-1)^{n-k}\binom{n}{k}\sum_{j=1}^{k}\frac{1}{j}\sum_{i=1}^{j}\frac{1}{i}.$$

Dividing by $(-1)^{n-1}$ and applying (1) once more yields the desired identity.

Solution III by O. P. Lossers, University of Technology, Eindhoven, The Netherlands. Let

$$f(x) = \sum_{k=1}^{n}\frac{(-1)^{k-1}}{k}\binom{n}{k}\sum_{j=1}^{k}\frac{1}{j}\sum_{i=1}^{k}\frac{1}{i}(1-x)^i.$$

We prove that $f(x) - f(0) = -\sum_{k=1}^{n} x^k/k^3$. Since $f(1) = 0$, this yields the desired identity $f(0) = \sum_{k=1}^{n} 1/k^3$.

Let Δ be the transformation taking $p(x)$ to $xp'(x)$. Direct computation shows that $\Delta^3(-\sum_{k=1}^{n} x^k/k^3) = -\sum_{k=1}^{n} x^k$. Since Δ is injective on the space of polynomials with constant term 0, it suffices to show that $\Delta^3\big(f(x) - f(0)\big) = -\sum_{k=1}^{n} x^k$ as well.

The linearity of Δ yields $\Delta^3\big(f(x) - f(0)\big) = \Delta^3\big(f(x)\big)$. Now we compute directly

$$\begin{aligned}\Delta^3\big(f(x)\big) &= \Delta^2\left(\sum_{k=1}^{n}\binom{n}{k}(-1)^k\frac{1}{k}\sum_{j=1}^{k}\frac{1-(1-x)^j}{j}\right)\\ &= \Delta\left(\sum_{k=1}^{n}\binom{n}{k}(-1)^k\frac{1-(1-x)^k}{k}\right)\\ &= x\sum_{k=1}^{n}\binom{n}{k}(-1)^k(1-x)^{k-1} = x\frac{x^n-1}{1-x} = -\sum_{k=1}^{n}x^k.\end{aligned}$$

Solution IV by Víctor Hernández, Universidad Nacional de Educación a Distancia, Madrid, Spain. The desired identity is the case $m = 3$ of the more general formula

$$\sum_{1\le i_1\le\cdots\le i_m\le n}\frac{(-1)^{i_m-1}}{i_1\cdots i_m}\binom{n}{i_m} = \sum_{k=1}^{n}\frac{1}{k^m}.$$

Consider two probabilistic processes for partitioning the unit interval $[0, 1]$ into $m+1$ subintervals. Let $X_0 = 1$ and $Y_0 = 0$. Choose X_{i+1} uniformly at random in $[0, X_i]$, and choose Y_{i+1} uniformly at random in $[Y_i, 1]$. By symmetry, X_m and $1 - Y_m$ have the same distribution. We compute

$$\mathbf{E}(X_m^{k-1}) = \int_0^1\frac{dx_1}{x_1}\int_0^{x_1}\frac{dx_2}{x_2}\cdots\int_0^{x_{m-1}}x_m^{k-1}\,dx_m = \frac{1}{k^m} \qquad \text{and}$$

$$\mathbf{E}(Y_m^{k-1}) = \int_0^1\frac{dy_1}{1-y_1}\int_{y_1}^1\frac{dy_2}{1-y_2}\cdots\int_{y_{m-1}}^1 y_m^{k-1}\,dy_m = \frac{1}{k}\sum_{i_{m-1}=1}^{k}\frac{1}{i_{m-1}}\cdots\sum_{i_1=1}^{i_2}\frac{1}{i_1}.$$

Thus

$$\begin{aligned}\sum_{k=1}^{n}\frac{1}{k^m} &= \mathbf{E}\left(\sum_{k=1}^{n}X_m^{k-1}\right) = \mathbf{E}\left(\frac{1-X_m^n}{1-X_m}\right) = \mathbf{E}\left(\frac{1-(1-Y_m)^n}{Y_m}\right)\\ &= \sum_{k=1}^{n}(-1)^{k-1}\binom{n}{k}\mathbf{E}\left(Y_m^{k-1}\right) = \sum_{1\le i_1\le\cdots\le i_m\le n}\frac{(-1)^{i_m-1}}{i_1\cdots i_m}\binom{n}{i_m},\end{aligned}$$

where we have set $i_m = k$.

Editorial comment. Other solvers used generating functions, inductive arguments, and various identities and transformations. Joe Howard and Heinz-Jürgen Seiffert independently observed that the result follows fairly quickly from a result proved by the proposer in his solution to Problem 4427, *School Science and Mathematics* 95 (1995) 221.

Solved also by E. S. Andersen & M. E. Larsen (Denmark), J. C. Binz (Switzerland), P. Bracken, D. Bradley (Canada), E. Braune (Austria), D. Callan, R. J. Chapman (U. K.), D. A. Darling, V. Dwivedi (India), E. Hertz, M. Hoffman, J. Howard, A. Kaplan (France), A. A. Kelzon (Russia), R. A. Kopas, O. Krafft & M. Schaefer (Germany), J. H. Lindsey II, J. Lorch, P. McCartney, D. K. Nester, A. Pechtl (Germany), F. Qi (China), E. Schmeichel, L. Scribani (South Africa), H.-J. Seiffert (Germany), A. Sinefakopoulos (Greece), I. Sofair, R. Sprugnoli (Italy), A. Stadler (Switzerland), A. Stenger, J. Van hamme (Belgium), M. Vowe (Switzerland), Z. Wu, GCHQ Problems Group (U. K.), WMC Problems group, and the proposer.

A Limit of Periods

10603 [1997, 567]. *Proposed by Yury J. Ionin and Robin R. Lewis, Central Michigan University, Mt. Pleasant, MI.* Let a, b, and k be positive integers, and let $P_k(a,b)$ be the period of the sequence $\{a^n \bmod b^k\}_{n=1}^{\infty}$. Find $\lim_{k\to\infty} P_{k+1}(a,b)/P_k(a,b)$.

Solution by the proposers. The limit equals the largest divisor of b that is relatively prime to a.

Suppose first that a and b are relatively prime. Fixing a and b, let $P_k = P_k(a,b)$. We have $a^{P_k} \equiv 1 \bmod b^k$, so $a^{P_k} = 1 + q_k b^k$ for some integer q_k. Note that P_k divides P_{k+1}. Thus $P_{k+1} = u_k P_k$, where u_k is the smallest positive integer u such that $a^{uP_k} \equiv 1 \bmod b^{k+1}$. Since

$$a^{uP_k} = (1+q_kb^k)^u \equiv (1+uq_kb^k) \bmod b^{k+1},$$

we have $u_k = b/d_k$, where $d_k = \gcd(q_k, b)$. Thus $P_{k+1} = bP_k/d_k$.

If $k \geq 2$, then the equalities $a^{P_{k+1}} = 1 + q_{k+1}b^{k+1} = (1+q_kb^k)^{b/d_k}$ imply that $q_{k+1} = (q_k/d_k) + t_kq_k^2b$, where t_k is an integer. Therefore, if p is a common prime divisor of b and q_k, then p occurs in the prime factorization of q_{k+1} with an exponent smaller than its exponent in the prime factorization of q_k. If p is a prime divisor of b that does not divide q_k, then p also does not divide q_{k+1}. If k is sufficiently large, this implies that q_k and b are relatively prime, so $d_k = 1$ and $P_{k+1} = bP_k$. It follows that the desired limit is b.

Suppose now that a and b are not relatively prime. Let r be the largest divisor of b that is relatively prime to a, and let $b = rs$. Now $P_k(a,b)$ is a period of the sequence $\{a^n \bmod r^k\}_{n=1}^{\infty}$, and thus $P_k(a,r)$ divides $P_k(a,b)$.

On the other hand, consider the expression $a^{P_k(a,r)+m} - a^m$. This is divisible by r^k, and for large m it is divisible by s^k. Since $\gcd(r,s) = 1$, it must therefore be divisible by b^k, and we discover that $P_k(a,r)$ is a period of the sequence $\{a^n \bmod b^k\}_{n=1}^{\infty}$. Hence $P_k(a,b)$ divides $P_k(a,r)$. We conclude that $P_k(a,b) = P_k(a,r)$. Since a and r are relatively prime, the claim follows from the previous case.

Solved also by D. A. Callan, R. J. Chapman (U. K.), J. H. Lindsey II, A. N. 't Woord (The Netherlands), GCHQ Problems Group, and NCCU Problems Group.

Avoiding the Identity

10606 [1997, 664]. *Proposed by Thomas Zaslavsky, Binghamton University, Binghamton, NY.* Given a positive integer m, show that there is a positive integer n such that, for every group G of order at least n, it is possible to choose m elements $g_1, g_2, \ldots, g_m$ so that no product of the form $g_{i_1}^{\pm 1} g_{i_2}^{\pm 1} \cdots g_{i_k}^{\pm 1}$ with $1 \leq k \leq m$ and distinct subscripts $i_1, i_2, \ldots, i_k$ in $\{1, 2, \ldots, m\}$ equals the identity.

Solution by Stephen M. Gagola, Jr., Kent State University, Kent, OH. Let a *signed product* be a product of distinct group elements or their inverses in a specified order. A set S is *admissible* if the identity is not expressible as a signed product of elements of S. We prove

that every group of order at least $f(m) = \lceil 2^{m-1}(m-1)!\sqrt{e}\,\rceil$ has an admissible subset of size m.

For $S \subseteq G$, let S^* be the set of signed products of elements of S. Let $g(m) = f(m+1)-2$. When $|S| = m$, we claim that $|S^*| \le g(m)$. A signed product is formed by choosing an ordered nonempty subset of S with exponents ± 1. Thus $|S^*| \le \sum_{k=1}^{m}\binom{m}{k}k!2^k$. We can rewrite the bound as $2^m m! T_{m-1}(1/2)$, where $T_{m-1}(x) = \sum_{j=0}^{m-1} x^j/j!$ is the Maclaurin polynomial of degree $m-1$ for e^x. The next term in the series expansion of $2^m m! e^{1/2}$ contributes 1, while the remainder after that is at most 1. Thus $|S^*| \le f(m+1)-2$.

Note that S^* and $G - S^*$ are closed under taking inverses. If a signed product equals the identity, then each of its elements can be expressed as a signed product of the other elements in the product. If S is admissible and x is a nonidentity element of $G - S^*$, it therefore follows that $S \cup \{x\}$ is also admissible. Thus S can be enlarged until $G - S^*$ contains only the identity element.

We now use induction on m to prove the claim that every group of order at least $f(m)$ has an admissible subset of size m. When $m = 1$, every nontrivial group has a nonidentity element, and this forms an admissible set of size 1. This agrees with $f(1) = 2$. When $m > 1$, the monotonicity of f and the induction hypothesis imply that every group of order at least $f(m)$ has an admissible subset S of size $m-1$. Since $g(m-1) = f(m)-2$, there is a nonidentity element in $G - S^*$, and S can be enlarged by one element.

Solved also by P. J. Anderson (Canada), R. Barbara (France), D. Beckwith, G. L. Body (U. K.), R. J. Chapman (U. K.), J. H. Lindsey II, S. C. Locke, J. Merickel, K. A. Ross, M. C. Slattery, GCHQ Problems Group (U. K.), NCCU Problems Group, NSA Problems Group, and the proposer.

Some Sums Require Care

10638 [1998, 69]. *Proposed by Brian Conolly, Cambridge, U. K.* For $0 \le \lambda \le 1$ and $m \ge 0$, let $S_m(\lambda) = \sum_{n\ge 1} e^{-\lambda n}(\lambda n)^{n-m}/n!$. Show that $S_0(\lambda) = \lambda/(1-\lambda)$, $S_1(\lambda) = 1$, $S_2(\lambda) = 1/\lambda - 1/2$, and $S_3(\lambda) = 1/\lambda^2 - 3/(4\lambda) + 1/6$.

Solution I by Allen Stenger, Tustin, CA. Let

$$T_m(\lambda) = \lambda^m S_m(\lambda) = \sum_{n\ge 1} \frac{(\lambda e^{-\lambda})^n n^{n-m}}{n!}.$$

By Stirling's formula the summand is asymptotic to $(2\pi)^{-1/2} n^{-m-1/2}(\lambda e^{1-\lambda})^n$. Thus, this sum converges absolutely for $|\lambda e^{1-\lambda}| < 1$. Therefore it represents a continuous function on $[0, 1)$. Furthermore, if $m \ge 1$, it converges uniformly for $|\lambda e^{1-\lambda}| \le 1$ and therefore is continuous on $[0, 1]$. For $m = 0$ it diverges at $\lambda = 1$.

First consider the case $m = 1$. We want to show that $\sum_{n\ge 1} (\lambda e^{-\lambda})^n n^{n-1}/n! = \lambda$. Euler showed that this holds for $0 \le \lambda \le 1$. It can be derived by applying the Lagrange inversion formula to $\lambda e^{-\lambda}$ (G. Pólya and G. Szegö, *Problems and Theorems in Analysis*, Volume 1, Springer, 1972, Part 3, Exercise 209).

The formulas for other values of m can be derived from the case $m = 1$ by integration or differentiation. Observe that $T_m(0) = 0$ and that by uniform convergence

$$\begin{aligned}\frac{d}{d\lambda} T_m(\lambda) &= \frac{d}{d\lambda} \sum_{n\ge 1} \frac{\lambda^n e^{-n\lambda} n^{n-m}}{n!} \\ &= \sum_{n\ge 1} \frac{\lambda^{n-1} e^{-n\lambda} n^{n-(m-1)}}{n!} - \sum_{n\ge 1} \frac{\lambda^n e^{-n\lambda} n^{n-(m-1)}}{n!} = \frac{1-\lambda}{\lambda} T_{m-1}(\lambda),\end{aligned}$$

for $\lambda \in [0, 1)$. Thus

$$T_0(\lambda) = \frac{\lambda}{1-\lambda}\frac{d}{d\lambda}T_1(\lambda) = \frac{\lambda}{1-\lambda},$$
$$T_2(\lambda) = \int_0^{\lambda} \frac{1-z}{z} T_1(z)\,dz = \lambda - \frac{1}{2}\lambda^2, \quad \text{and}$$
$$T_3(\lambda) = \int_0^{\lambda} \frac{1-z}{z} T_2(z)\,dz = \lambda - \frac{3}{4}\lambda^2 + \frac{1}{6}\lambda^3$$

for $0 \le \lambda < 1$. Continuity extends the latter two formulas to $\lambda = 1$.

Solution II by Thomas Hermann, SDRC, Milford, OH. Replacing $e^{-\lambda n}$ by its Taylor series and rearranging the sum formally gives

$$S_m(\lambda) = \sum_{n=1}^{\infty} \frac{(\lambda n)^{n-m}}{n!} \sum_{k=0}^{\infty} (-1)^k \frac{(\lambda n)^k}{k!} = \sum_{r=1}^{\infty} \frac{(-1)^r \lambda^{r-m}}{r!} \sum_{n=1}^{r} (-1)^n \binom{r}{n} n^{r-m}. \qquad (*)$$

This reordering of the sum is justified as follows. By the uniform convergence proved in Solution I, the first sum in $(*)$ represents an analytic function on the domain $\{\lambda : |\lambda e^{1-\lambda}| < 1\}$. For sufficiently small $|\lambda|$, this sum converges absolutely. Thus, the reordering is valid for $|\lambda|$ small. Hence the second sum in $(*)$ is the Laurent series for the analytic function given by the first sum in some punctured neighborhood of 0. We show that this Laurent series converges for $0 < |\lambda| < 1$ if $m = 0$ and for $\lambda \neq 0$ if $m \geq 1$. Therefore these two analytic functions agree on the connected domain $\{\lambda : |\lambda e^{1-\lambda}| < 1 \text{ and } 0 < |\lambda| < 1\}$ and on its boundary if $m \geq 1$. Thus, the two sides of $(*)$ agree for $\lambda \in [0, 1)$, and if $m \geq 1$ for $\lambda \in [0, 1]$.

The inner sum on the right side of $(*)$ is found in I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, 1994, equations 0.154 (3) and (4). It can also be derived by evaluating at 0 the $(r-m)$th derivative of $(e^t - 1)^r$ in two different ways. One obtains

$$\sum_{n=1}^{r} (-1)^n \binom{r}{n} n^{r-m} = \begin{cases} 0 & \text{if } 1 \leq m < r; \\ (-1)^r r! & \text{if } m = 0. \end{cases}$$

Therefore

$$S_0(\lambda) = \sum_{r=1}^{\infty} \frac{(-\lambda)^r}{r!}(-1)^r r! = \frac{\lambda}{1-\lambda},$$

which holds for $|\lambda| < 1$, and for $m \geq 1$,

$$S_m(\lambda) = \sum_{r=1}^{m} \left(\sum_{n=1}^{r} (-1)^{r-n} \frac{n^{r-m}}{n!(r-n)!} \right) \left(\frac{1}{\lambda} \right)^{m-r},$$

which holds for all $\lambda \neq 0$.

Editorial comment. The proposer notes that the problem of showing $S_1(1) = 1$ was posed by P. J. Cameron in connection with a result of Rényi concerning random forests of rooted trees. In queuing theory, $S_1(\lambda)$ is the probability that a busy period will ever end in the process M/D/1 with traffic intensity λ. Since $S_1(\lambda) < 1$ for $\lambda > 1$, a busy period for that process has nonzero probability of continuing forever. When $0 < \lambda < 1$, the mean number of customers served during a busy period is $S_0(\lambda)/\lambda$.

Solved also by P. Bracken (Canada), D. Callan, R. J. Chapman (U. K.), G. L. Isaacs, J. H. Lindsey II, O. P. Lossers (The Netherlands), V. Lucic (Canada), R. Mortini (France), V. Schindler (Germany), H.-J. Seiffert (Germany), A. Sofo (Australia), A. Stadler (Switzerland), A. Tissier (France), T. V. Trif (Romania), J. Van hamme (Belgium), M. Vowe (Switzerland), WMC Problems Group, and the proposer.

REVIEWS

Edited by **Harold P. Boas**
Mathematics Department, Texas A & M University, College Station, TX 77843-3368

An Imaginary Tale: The Story of $\sqrt{-1}$. By Paul J. Nahin. Princeton University Press, 1998, xvi + 257 pp., $24.95.

Reviewed by **Ricardo Diaz**

An Imaginary Tale is a leisurely, entertaining, idiosyncratic account of the development of the complex number system that can be enjoyed both by novices who know very little about complex numbers and by experts who think they know almost everything about them. The author, a professor of electrical engineering who also writes science fiction, has a special talent for expressing his infatuation with this fascinating topic. This accomplished raconteur has artfully avoided a direct chronological presentation of his story. He flouts the tedious conventions of linear time to successfully weave a complicated tapestry that combines historical anecdotes, doggerel, mathematical puzzles, fascinating calculations, the occasional theorem, and even samples of Cauchy's love letters.

In the first half of the book, you will learn how close the Babylonians came to inventing i, discover the name of the Norwegian surveyor who gave complex arithmetic its first geometrical description (a decade before Argand), and marvel at the discomfort many distinguished mathematicians (including Leibniz, Boole, and Airy) suffered in the presence of the imaginary. If you proceed to the second half, you will be exposed to a crash course in contour integration and the Cauchy-Riemann equations, plus a sampling of applications in engineering and pure science. You will also see some beautiful gems from pure mathematics, such as Euler's heuristic argument for expressing the sine function as an infinite product, culminating in his derivation of the identity $\pi^2/6 = \sum_{n=1}^{\infty} n^{-2}$. As a bonus, you will find in the appendices a discussion of how i^i was once computed to 135 decimals in 1921 using tables of logarithms.

Of course this book is not a textbook, nor a scholarly treatise on the history of mathematics, and we must not judge it by pedantic standards. The aim of this book is far more ambitious than that of a textbook intended to instruct: this book was written to inspire and delight. Mathematics instructors at the college level will savor the amusing anecdotes and pass on choice items to colleagues and students. Eager undergraduates will enjoy learning that great mathematicians struggled over ideas that are now taken as obvious, and bright high school students will discover the adventures that await them if they persevere in their mathematical education.

The enthusiastic author sometimes ascends to heights that most of his intended audience will be unable to follow, as in his discussion of the functional equation for the Riemann zeta function. Setting such passages aside, there is still a tremendous quantity of simple but elegant mathematics to enthrall, amuse, edify, surprise, and enchant even the most jaded of mathematical sophisticates. This is

indeed a rare book: one that lives up to the self-praise that adorns its jacket. Professor Nahin has delivered the real goods.

The reader who wishes to explore the history of the complex number system in greater depth may also wish to consult another informative, albeit less whimsical, book: *Numbers*, Heinz-Dieter Ebbinghaus et al., Springer, 1991.

University of Northern Colorado, Greeley, CO 80639
rdiaz@bentley.unco.edu

Leaning Towards Infinity. By Sue Woolfe. Faber and Faber, Boston, 1997 (originally published by Random House Australia, 1996), xxi + 393 pp., $24.95 hardcover, $14.95 softcover.

Reviewed by **John Beebee** and **Karen Willmore**

Are you curious about why men outnumber women in your advanced mathematics classes? Or are you interested in what a gifted writer who claims to know nothing of mathematics thinks about the feelings and motivations of its practitioners? If so, then you will find this a tempting novel, as we did. Its premise is that the search for truth in life, love, and mathematics is a messy and risky business.

> I think it all began because of the shape of my mother's breasts. And it definitely began with something my mother wrote on the margin of a page stuck on the wall: Frege said that the line connecting any two points is already there before we draw it. (p. 3)

What is this "it" described by Frances Montrose, a gifted amateur mathematician, damaged daughter of another gifted amateur, Juanita, and mother of her biographer, Hypatia? (And what if the Bernoullis had been mothers and daughters instead of fathers and sons?) Frances has just won a competition for amateur mathematicians and has the opportunity to present the joint ideas of herself and her mother at a conference in Athens, Greece, the birthplace of modern mathematics. In another possible beginning, Hypatia summarizes the plot as the story of "an Australian woman who'd never had any formal mathematics training but who in 1995 carried across the world, inside a borrowed suitcase and amongst ball dresses, a bulging three hundred and fifteen pages of revolutionary theorems, and something else no one knew about—the beginnings of the discovery of a new kind of number" (p. xiii).

The emotional core of the novel is interwoven around four generations of unresolved mother-daughter relationships. Beautiful Juanita, abandoned by her self-centered mother, is a mathematical genius who in unthinking rebellion marries a man who can't count past ten. Her mathematics, compulsively pursued during stolen time, almost reaches a breakthrough. She pins her love and her mathematical dreams on her beautiful son, who has no use for either. Frances, her very plain daughter, is so consumed with trying to win her mother's love and recognition by completing her mother's mathematics that she neglects her own daughter, Hypatia

—who despite her name, claims that she can barely add. Hypatia tries to break this destructive cycle with her own daughter, Zoe, by writing down the story of her mother's life. Is it possible, this novel asks, to be both a mother and a great person? For as Hypatia writes, motherhood is "about the allocation of the soul" (p. 216).

In an interview on Wisconsin Public Radio (The Beauty of Math, 11/4/97), Sue Woolfe told of how she, a nonmathematician, came to use mathematics as a metaphor of who we are. She was living in an isolated place, nursing a baby, and listening to the ramblings of her neighbor, a mathematician dying of cancer. He talked not about his disease but about mathematics, and she came to realize that mathematics was as exotic and wonderful to him as her baby was to her. Woolfe went on to explain how she sees the connection between mathematics and love: both are full of assumptions that the mathematician and the lover accept on faith, with no guarantees of certainty.

According to the root meaning of the word, amateurs are those who love what they do, and it is in this sense that Woolfe characterizes her very gifted female mathematicians as amateurs. In a family tree at the beginning of the novel, Juanita is labeled a "housewife and secret mathematician," Frances an "English teacher and amateur mathematician," and her one-time husband a "professional mathematician." Frances' husband would never discuss mathematics with her and left her as soon as their daughter was born. Woolfe portrays professional mathematicians mostly as conceited, ambitious, and rude (and male), more concerned with defending the established profession against outsiders than with mathematical ideas. She comes down solidly on the side of "cranks," those rare brilliant outsiders who are in love with mathematics. She even begins her novel (yes, another beginning) with the most famous crank story of all—the story of Srinivasa Ramanujan.

Juanita once confesses to a young Frances, "I'm in love with thinking. I don't mean thinking things through or thinking things over or thinking things out.... Thinking itself.... It's as abstract as death.... All the mess of life removed. And what's left, it can go anywhere" (pp. 122–123). For Frances, mathematics is very much connected with her mother's beauty: "She's always been mathematics to me, beautiful beyond compare, unbearably elusive, unbearably unknowable" (p. 250). She says doing mathematics is like being "in an undiscovered country... and I'm like a blind person. There's something up ahead of me, and I don't know what it is. I'm feeling with my hands a bit sticking out here, and a bit going in there. Maybe over here is an arch. Maybe over there is a tunnel. It's like something that might be a labyrinthine cathedral" (p. 46).

Woolfe tells the story of these women's lives and loves in anything but a straightforward, linear fashion. She warns us of her casual attitude toward time in the family tree, which ends with Juanita's great grandson being born in 2003. Frances relates her experiences in Greece as occurring in the present, but then all through her story (which actually is being told by Hypatia as if she were her mother) she randomly digresses to recall her childhood, which in turn is interspersed with recalling her mother's story. Hypatia, meanwhile, constantly interrupts all this retelling with her own asides, sometimes interpreting events, sometimes telling her own story as Frances' daughter, sometimes telling stories of famous mathematicians from the past (including the historical Hypatia). Initially, this is all very confusing for us as readers, and a certain amount of perseverance and faith is required before one begins to sort out just what is happening to whom and when—which may be not unlike doing mathematics. At any rate, Hypatia, in one of her digressions, gives this justification: "My mother's mind was probably

never one of those minds that travel in straight lines. Who knows? That may have been the secret of her greatness.... (Some might say that digression is to be expected of a mathematician, that mathematics *is* a digression)" (pp. 9–10).

Before assigning *Leaning Towards Infinity* to our book club, composed of mostly academic types with a token physician, I (Karen) had John read it to make sure that the mathematics made sense. All but one book club member liked the novel very much; all agreed that it was confusing at first; one person disliked the metaphoric and image-laden language. Australians really like this book: it was the best-selling Australian novel in 1996.

I (John) half anticipated a mixture of feminism and superficial impressions of what mathematicians do, written in a painful and pretentious literary style. What I found was different. Although the prose was painful to me at first, it became more natural as I got drawn in and began to untangle the threads of the story. Woolfe has an unusually sensitive grasp of the process of mathematical creation, and I found her thinking about motherhood and mathematics to be uniquely her own. The mathematics conference in Athens—surreal, funny, and moving—should not be missed.

Department of Mathematical Sciences, University of Alaska Anchorage, Anchorage, AK 99508
afjcb@uaa.alaska.edu
Women's Studies Program, University of Alaska Anchorage, Anchorage AK 99508
afkew@uaa.alaska.edu

From the MONTHLY 25 Years Ago...

THE COMPUTER ILLITERACY PROBLEM: A PARTIAL SOLUTION

PETER G. LYKOS

1. How the computer is affecting education. The computer is transforming the ways in which the problem solvers and the decision makers of our society go about doing their jobs. Not only are old problem-solving techniques being speeded up and scaled up, but new techniques are being invented and developed which would not have been considered seriously before the invention and proliferation of the computer. And the proliferation of the computer is being accelerated as a direct consequence of two technological developments:

1. The low cost and increasing flexibility of *minicomputers*, and
2. The increasing flexibility and ease of use of *tele-communications* systems whereby users can use a typewriter-like or keyboard plus TV-like terminal to access a variety of computers remotely located.

MONTHLY 81 (1974) 393

TELEGRAPHIC REVIEWS

Edited by **Arnold Ostebee**
with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges

Telegraphic Reviews are designed to alert readers in a timely manner to new books appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T : Textbook	*P : Professional Reading*	*1–4: Semester*
C : Computer Software	*L : Undergraduate Library*	*** : Special Emphasis*
S : Supplementary Reading	*13: Grade Level*	*?? : Questionable*

Readers are advised that price information is subject to change. Selected books receive a second, more extensive review in the *Monthly*.

Books submitted for review should be sent to *Book Reviews Editor, American Mathematical Monthly, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098.*

Precalculus, T(13: 1), C. *Explorations in College Algebra.* Linda Almgren Kime, Judy Clark. Wiley, 1998, xviii + 649 pp, $73.95 (P), with CD ROM. [ISBN 0-471-10698-4] Designed "to shift the focus from learning a set of discrete mathematical rules to exploring how algebra is used to answer questions about the physical and social world around us." Similar in spirit to other reform texts; reflects NCTM and AMATYC standards.

Precalculus, S(13). *Maths: A Student's Survival Guide.* Jenny Olive. Cambridge Univ Pr, 1998, xiii + 564 pp, $29.95 (P); $74.95. [ISBN 0-521-57586-9; 0-521-57306-8] Self-help workbook format; over 800 questions with detailed solutions. Topics range from basic algebra through single-variable calculus. Aimed at science and engineering students.

Foundations, S(14–16: 1), S, P, L. *Explaining Chaos.* Peter Smith. Cambridge Univ Pr, 1998, viii + 193 pp, $59.95; $19.95 (P). [ISBN 0-521-471710; 0-521-477476] Philosophical treatment of chaos theory—its foundations, history, and modern role in science. Good, accessible explanations, nice examples, some rigorous mathematical descriptions. RM

Combinatorics, P*, L. *Handbook of Combinatorial Optimization, Volumes 1–3.* Eds: Ding-Zhu Du, Panos M. Pardalos. Kluwer Academic, 1998, $1325 set, [ISBN 0-7923-5019-7] set. *Volume 1*, viii + 785 pp; *Volume 2*, viii + 753 pp; *Volume 3*, viii + 865 pp. 33 expository articles survey algorithmic approaches to many discrete and combinatorial problems. Most articles are accessible to nonspecialists.

Discrete Mathematics, T(13–14: 1), S, L*. *Mathematical Problems and Proofs: Combinatorics, Number Theory, and Geometry.* Branislav Kisačanin. Plenum Pr, 1998, xiv + 220 pp, $55. [ISBN 0-306-45967-1] Introduction to basic results and techniques. Emphasizes interesting examples and problems. DB

Number Theory, P. *Random Matrices, Frobenius Eigenvalues, and Monodromy.* Nicholas M. Katz, Peter Sarnak. Colloquium Public., V. 45. AMS, 1999, xi + 419 pp, $69. [ISBN 0-8218-1017-0] There is empirical evidence that the distribution of the spacings between zeroes of the zeta-function is the same as a certain probability measure from random matrix theory. This book establishes this relationship for several classes of zeta- and L-functions over finite fields. DB

Algebra, P. *Elimination Methods in Polynomial Computer Algebra.* Valery Bykov, *et al.* Math. & Its Applic., V. 448. Kluwer Academic, 1998, xi + 237 pp, $106. [ISBN 0-7923-5240-8] Methods based on multidimensional residue theory. Illustrates theory with applications to mathematical kinetics.

Algebra, P. *Higher Category Theory.* Eds: Ezra Getzler, Mikhail Kapranov. Contemp. Math., V. 230. AMS, 1998, x + 134 pp, $34 (P). [ISBN 0-8218-1056-1] Proceedings of a 1997 workshop at Northwestern University.

Algebra, P. *Trends in the Representation Theory of Finite Dimensional Algebras.* Eds: Edward L. Green, Birge Huisgen-Zimmermann. Contemp. Math., V. 229. AMS, 1998, xiii + 356 pp, $75 (P). [ISBN 0-8218-0928-8] Pro-

ceedings of the 1997 Joint Summer Research Conference at the University of Washington.

Calculus, S(13), L.** *Calculus Mysteries and Thrillers.* R. Grant Woods. MAA, 1998, xix + 131 pp, $24.95 (P). [ISBN 0-88385-711-1] 11 projects, each in the form of an amusing short story, designed to develop students' modeling and technical writing skills. Appropriate for use in single-variable calculus courses. Includes a sample solution for each project. AO

Calculus, T(13: 1). *Workshop Calculus: Guided Exploration with Review, Volume 2.* Nancy Baxter Hastings, *et al.* Springer-Verlag, 1998, xxiv + 397 pp, $39.95 (P). [ISBN 0-387-98349-X] Second of a two-volume series that integrates review of precalculus topics with standard first-semester calculus topics. Designed for use in a "workshop" classroom environment, each volume is a collection of guided exploration activities and homework exercises. (*Volume 1*, TR, October 1997.)

Calculus, T(14: 1). *Vector Calculus.* P.C. Matthews. Springer-Verlag, 1998, ix + 182 pp, $29.95 (P). [ISBN 3-540-76180-2] Concise introduction. Many references to topics in the physical sciences. Uses suffix notation; includes a chapter on Cartesian tensors. Assumes some knowledge of linear algebra (matrices and determinants), and basic calculus (differentiation, integration, and partial differentiation). Includes solutions to all of the exercises. AO

Calculus, T(13: 3). *Calculus: A New Horizon, Sixth Edition.* Howard Anton. Wiley, 1999, xxxii + 1314 pp, $94.95. [ISBN 0-471-15306-0] Major revision. More emphasis on conceptual understanding and applicability; some reorganization of topics. (*Fifth Edition*, TR, October 1995.)

Complex Analysis, T(18), S, P. *The Logarithmic Integral I.* Paul Koosis. Stud. in Adv. Math., V. 12. Cambridge Univ Pr, 1998, xviii + 606 pp, $47.95 (P); $130. [ISBN 0-521-59672-6; 0-521-30906-9] Corrected paperback edition. (*First Edition*, TR, May 1989.)

Differential Equations, P. *Differential Equations: La Pietra 1996.* Eds: M. Giaquinta, J. Shatah, S.R.S. Varadhan. Proc. of Symp. in Pure Math., V. 65. AMS, 1999, xi + 219 pp, $39. [ISBN 0-8218-0610-6] Proceedings of a conference held to honor the 70th birthdays of Peter Lax and Louis Nirenberg.

Differential Equations, T(14: 1). *Differential Equations: Graphics, Models, Data.* David Lomen, David Lovelock. Wiley, 1999, xiv + 682 pp, $96.95. [ISBN 0-471-07648-1] Covers standard set of topics using numerical, graphical, analytical, and descriptive viewpoints. Encourages an exploratory approach. Motivates many topics using real-world problems or data. AO

Partial Differential Equations, P. *Parametric Lie Group Actions on Global Generalised Solutions of Nonlinear PDEs.* Elemér E. Rosinger. Math. & Its Applic., V. 452. Kluwer Academic, 1998, xvii + 234 pp, $106. [ISBN 0-7923-5232-7]

Partial Differential Equations, T(15: 1), L. *Beginning Partial Differential Equations.* Peter V. O'Neil. Wiley, 1999, x + 500 pp, $79.95. [ISBN 0-471-23887-2] For a first course in PDEs. Topics: method of characteristics; classification of second-order equations; well-posedness; Fourier series; wave and heat equations; Dirichlet and Neumann problems. Some exercises require use of a CAS. PG

Numerical Analysis, P. *Parameter Estimation in Nonlinear Dynamical Systems.* W.J.H. Stortelder. CWI Tract, V. 124. Centrum voor Wiskunde en Informatica, 1998, vi + 176 pp, Dfl. 40 (P). [ISBN 90-6196-482-2] Numerical and statistical aspects of the parameter estimation problem for dynamical systems described by differential algebraic equations. Includes several applied case studies.

Numerical Analysis, S(16–17), P, L*. *Matrix Algorithms, Volume I: Basic Decompositions.* G.W. Stewart. SIAM, 1998, xix + 458 pp, $32 (P). [ISBN 0-89871-414-1] First of a planned 5-volume series for non-specialists. Emphasizes algorithms, their derivation, and their analysis. Initial chapters provide basic background on matrices, linear algebra, and the realities of matrix computations on computers. Subsequent chapters discuss LU, QR, and rank-reducing decompositions. AO

Operator Theory, P. *Operator Algebras and Operator Theory.* Eds: Liming Ge, *et al.* Contemp. Math., V. 228. AMS, 1998, xx + 389 pp, $85 (P). [ISBN 0-8218-1093-6] Proceedings of a 1997 conference in Shanghai, China.

Operator Theory, P. *Introduction to Vertex Operator Superalgebras and Their Modules.* Xiaoping Xu. Math. & Its Applic., V. 456. Kluwer Academic, 1998, xvi + 356 pp, $159. [ISBN 0-7923-5242-4]

Functional Analysis, P. *Geometría de Espacios de Banach.* José A. Facenda Aguirre. Universidad de Sevilla (Porvenir, 27—41013 Sevilla, SPAIN), 1998, vi + 139 pp, (P). [ISBN 84-472-0473-1]

Analysis, S(18), P. *Advanced Integration Theory.* Corneliu Constantinescu, *et al.* Math. &

Its Applic., V. 454. Kluwer Academic, 1998, x + 861 pp, $375. [ISBN 0-7923-5234-3] This huge work aims to tie together theories used by probabilists and analysts. Vector lattice theory provides the framework for their study, including L^p-spaces, real measures, absolute continuity. Each section contains many exercises. Good for self-study. KS

Algebraic Geometry, P. *Algebra and Geometry.* Ed: Ming-chang Kang, *et al.* International Pr, 1998, ix + 227 pp. [ISBN 1-57146-058-6] Proceedings of a 1995 conference at National Taiwan University.

Differential Geometry, P. *The Theory of Finslerian Laplacians and Applications.* Eds: Peter L. Antonelli, Bradley C. Lackey. Math. & Its Applic., V. 459. Kluwer Academic, 1998, xxx + 282 pp, $146. [ISBN 0-7923-5313-7]

Differential Geometry, P. *Harmonic Maps, Loop Groups, and Integrable Systems.* Martin A. Guest. London Math. Soc. Stud. Texts, V. 38. Cambridge Univ Pr, 1997, xiv + 194 pp, $21.95 (P); $59.95. [ISBN 0-521-58932-0; 0-521-58085-4] "The specific goal of this book is to show how the theory of loop groups can be used to study harmonic maps."

Differential Geometry, P. *New Developments in Differential Geometry, Budapest 1996.* Ed: J. Szenthe. Kluwer Academic, 1999, xii + 519 pp, $227. [ISBN 0-7923-5307-2] 36 papers based on presentations at the conference.

Differential Geometry, P. *Topics in Symplectic 4-Manifolds.* Ed: Ronald J. Stern. Lect. Ser., V. 1. International Pr, 1998, iii + 124 pp. [ISBN 1-57146-019-5] 6 papers based on invited lectures presented at a 1996 conference at the University of California at Irvine.

Geometry, P. *Monge Ampère Equation: Applications to Geometry and Optimization.* Eds: Luis A. Caffarelli, Mario Milman. Contemp. Math., V. 226. AMS, 1999, ix + 172 pp, $39 (P). [ISBN 0-8218-0917-2] 10 invited lectures and contributed papers from a 1997 NSF–CBMS conference held at Florida Atlantic University.

Topology, P. *Continuous Selections of Multivalued Mappings.* Dušan Repovš, Pavel Vladimirovič Semenov. Math. & Its Applic., V. 455. Kluwer Academic, 1998, viii + 356 pp, $159. [ISBN 0-7923-5277-7]

Operations Research, P. *Fuzzy Sets in Decision Analysis, Operations Research and Statistics.* Ed: Roman Słowiński. Handbooks of Fuzzy Sets Ser. Kluwer Academic, 1998, xxiv + 453 pp, $169.95. [ISBN 0-7923-8112-2] 13 chapters in four sections: Decision Making; Mathematical Programming; Statistics and Data Analysis; Reliability, Maintenance, and Replacement.

Game Theory, T(18), S, P. *Dynamic Noncooperative Game Theory, Second Edition.* Tamer Başar, Geert Jan Olsder. Classics in Appl. Math., V. 23. SIAM, 1999, xv + 519 pp, $49.50 (P). [ISBN 0-89871-429-X] Unabridged, revised republication of the 1995 *Second Edition* originally published by Academic Press. This edition includes new material on the Braess Paradox and on the relationship between the existence of solutions of Ricatti equations and the existence of Nash equilibrium solutions. (*First Edition*, TR, April 1984.)

Mathematical Modeling, T(15), L. *Mathematical Analysis for Modeling.* Judah Rosenblatt, Stoughton Bell. Math. Modeling Ser. CRC Pr, 1999, 860 pp, $69.95. [ISBN 0-8493-8337-4] Presents mathematical framework for modeling problems in science and technology. Topics covered include Riemann and Lebesgue integration, Taylor's theorem, infinite series, multivariable calculus, coordinate systems, matrices, Fourier transforms, and generalized functions. PG

Optimal Control, P. *Differential Geometry and Control.* Eds: G. Ferreyra, *et al.* Proc. of Symp. in Pure Math., V. 64. AMS, 1999, viii + 341 pp, $79. [ISBN 0-8218-0887-7] 20 papers from the 1997 AMS Summer Research Institute held at the University of Colorado, Boulder.

Stochastic Processes, P. *Martingales et chaînes de Markov.* Laurent Mazliak, Pierre Priouret, Paolo Baldi. Hermann, 1998, viii + 215 pp, 180 F (P). [ISBN 2-7056-6382-7]

Stochastic Processes, P. *Optimal Filtering, Volume I: Filtering of Stochastic Processes.* Vladimir Fomin. Math. & Its Applic., V. 457. Kluwer Academic, 1999, xiii + 375 pp, $173. [ISBN 0-7923-5286-6]

Stochastic Processes, P. *Stochastic Partial Differential Equations: Six Perspectives.* Eds: Rene A. Carmona, Boris Rozovskii. Math. Surv. & Mono., V. 64. AMS, 1999, xi + 334 pp, $49. [ISBN 0-8218-0806-0] 6 survey articles; each presents a different perspective on the primary topic.

Stochastic Processes, P. *One-Dimensional Random Polymers.* R.W. van der Hofstad. CWI Tract, V. 123. Centrum voor Wiskunde en Informatica, 1998, 165 pp, Dfl. 40 (P). [ISBN 90-6196-481-4]

Statistical Methods, P. *Maximum Entropy and Bayesian Methods.* Eds: Gary J. Erickson, Joshua T. Rychert, C. Ray Smith. Fund.

Theories of Physics, V. 98. Kluwer Academic, 1998, ix + 297 pp, $133. [ISBN 0-7923-5047-2] Proceedings of the 17th International Workshop held at Boise State University in 1997.

Algorithms, T(15–17: 1), S, P, L. *Combinatorial Algorithms: Generation, Enumeration, and Search.* Donald L. Kreher, Douglas R. Stinson. Disc. Math. & Its Applic. CRC Pr, 1999, 329 pp, $74.95. [ISBN 0-8493-3988-X] Crisp modern treatment. Standard topics plus generation of combinatorial objects, groups and symmetry, computing isomorphisms, basis reduction. Nice concrete examples. RM

Computer Science, T(14–15: 1). *Computer Systems.* J. Stanley Warford. Jones & Bartlett, 1999, xv + 506 pp. [ISBN 0-7637-0794-5]

Computer Science, P. *Reliable Computer Systems: Design and Evaluation, Third Edition.* Daniel P. Siewiorek, Robert S. Swarz. AK Peters, 1998, xix + 908 pp, $65. [ISBN 1-56881-092-X]

Applications (Biological Science), T(16–17: 2), P, L*. *Mathematical Physiology.* James Keener, James Sneyd. Interdisc. Appl. Math., V. 8. Springer-Verlag, 1998, xix + 766 pp, $69.95. [ISBN 0-387-98381-3] An introductory survey emphasizing continuous, deterministic approaches. Illustrates how mathematics provides insight into physiological questions as well as how physiological questions can lead to new mathematical problems. First part presents the fundamentals of cellular physiology; second part discusses the physiology of systems. Accessible to mathematicians with little knowledge of physiology. AO

Applications (Engineering), T(14–15: 4). *Advanced Engineering Mathematics, Eighth Edition.* Erwin Kreyszig. Wiley, 1999, xvi + 1273 pp, $109.95. [ISBN 0-471-15496-2] New edition of this classic text/reference (*Seventh Edition*, TR, January 1995). Major changes: more emphasis on qualitative methods and applications in problem sets; projects of various types (e.g., team, writing, CAS); updated chapters on numerical methods. AO

Applications (Fluid Mechanics), P. *Applications of Group-Theoretical Methods in Hydrodynamics.* V.K. Andreev, *et al.* Math. & Its Applic., V. 450. Kluwer Academic, 1998, xii + 396 pp, $205. [ISBN 0-7923-5215-7]

Applications (Fluid Mechanics), P. *Theory of Multicomponent Fluids.* Donald A. Drew, Stephen L. Passman. Appl. Math. Sci., V. 135. Springer-Verlag, 1999, x + 308 pp, $59.95. [ISBN 0-387-98380-5]

Applications (Physics), T(16–17). *Classical Electrodynamics.* Walter Greiner. Theor. Physics. Springer-Verlag, 1998, x + 555 pp, $49.95 (P). [ISBN 0-387-94799-X] Careful presentation of classical electrostatics, magnetostatics, and electrodynamics. No exercises, but many worked examples. AO

Applications (Quantum Theory), P. *Quantum Measures and Spaces.* G. Kalmbach. Math. & Its Applic., V. 453. Kluwer Academic, 1998, xi + 343 pp, $159. [ISBN 0-7923-5288-2]

Applications (Quantum Theory), P. *Dirac Operators and Spectral Geometry.* Giampiero Esposito. Lect. Notes in Phys., V. 12. Cambridge Univ Pr, 1998, xiii + 209 pp, $34.95 (P). [ISBN 0-521-64862-9]

Applications (Systems Theory), P. *Stability of Finite and Infinite Dimensional Systems.* Michael I. Gil'. Intern. Ser. in Engin. & Comp. Sci. Kluwer Academic, 1998, xviii + 358 pp, $145. [ISBN 0-7923-8221-8]

Applications (Systems Theory), P. *Lecture Notes in Control and Information Sciences–241: Learning, Control and Hybrid Systems.* Eds: Yutaka Yamamoto, Shinji Hara. Springer-Verlag, 1999, xv + 451 pp, $92 (P). [ISBN 1-85233-076-7] Proceedings of a 1998 workshop in Bangalore, India, honoring B.A. Francis and M. Vidyasagar.

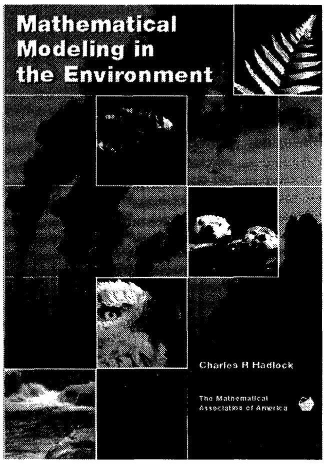
Applications (Systems Theory), P. *Lecture Notes in Control and Information Sciences–240: Low Gain Feedback.* Zongli Lin. Springer-Verlag, 1999, xvii + 354 pp, $105 (P). [ISBN 1-85233-081-3]

Applications (Systems Theory), P. *Lecture Notes in Control and Information Sciences–238: Computational Methods for Controller Design.* Nicola Elia, Munther A. Dahleh. Springer-Verlag, 1998, xi + 153 pp, $59 (P). [ISBN 1-85233-075-9]

Applications (Systems Theory), P. *Lecture Notes in Control and Information Sciences–239: Finite Spectrum Assignment for Time-Delay Systems.* Qing-Gou Wang, Tong Heng Lee, Kok Kiong Tan. Springer-Verlag, 1999, viii + 117 pp, $45 (P). [ISBN 1-85233-065-1]

Reviewers

DB: David Bressoud, Macalester; PG: Philip Gloor, St. Olaf; RM: Richard Molnar, Macalester; AO: Arnold Ostebee, St. Olaf; KS: Karen Saxe, Macalester; PZ: Paul Zorn, St. Olaf.



THE MATHEMATICAL ASSOCIATION OF AMERICA

Mathematical Modeling in the Environment

Charles Hadlock

Series: Classroom Resource Materials

Packaged with a PC compatible disk that enhances the material in the text.

Suitable for classroom adoption in an innovative course for

- **a general education mathematics elective**
- **a mathematics or science major advanced elective**
- **an interdisciplinary course, even at a relatively elementary level**
- **a mathematical modeling course in a civil/environmental engineering program**

This book has a dual objective: first, to introduce the reader to some of the most important and widespread environmental issues of the day; and second, to illustrate the vital role played by mathematical models in investigating these issues. The environmental issues addressed include: ground-water contamination, air pollution, and hazardous material emergencies. These issues are presented in their full real-world context, not as scientific or mathematical abstractions; and for background, readers are invited to investigate their status in their own communities.

The first part of the book leads the reader through relatively elementary modeling of these phenomena, including simple algebraic equations for ground water, slightly more complex algebraic equations (preferably implemented on a spreadsheet or other computerized framework) for air pollution, and a fully computerized modeling package for hazardous materials incident analysis. The interplay between physical intuition and mathematical analysis is emphasized.

For more advanced readers, the second part of the book returns to the same three subjects but with a higher level of mathematical sophistication (adjustable to the preparation of the reader by selection of subsections.) Many important classical mathematical themes are developed through this context, examples coming from single and multivariable calculus, differential equations, numerical analysis, linear algebra and probability. The material is presented in such a way as to minimize the required background and to encourage the subsequent study of some of these fields.

An elementary course for a general audience could be based entirely on Part I, and a higher level mathematics, science, or engineering course could move quickly to Part 2.

A PC compatible diskette packaged with the text contains a spreadsheet program that facilitates the numerical experimentation with the Gaussian plume equation introduced in Chapter 3, as well as public domain DOS program (ARCHIE) for evaluating the consequences from various hazardous materials scenarios (e.g., the physical extent of flammable and toxic vapor clouds). Text is not tied to the use of this software, but it is included as an aid to meet the pedagogical objectives of the text.

Catalog Code: ENV/SA
312 pp., Paperbound, 1998, ISBN 0-88385-709-X
List: $55.00 MAA Member: $43.95

Instructor's and Solutions Manual for *Mathematical Modeling in the Environment*

Charles Hadlock

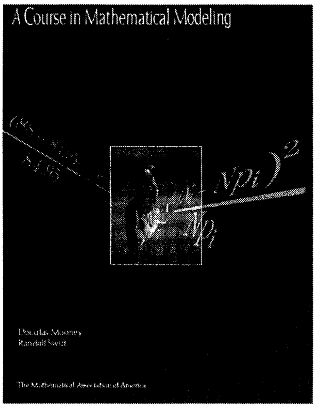
Contains the complete solutions and further discussion of nearly every exercise presented in the textbook. This includes both the mathematical/computational exercises as well as the research questions and investigations. Readers will benefit greatly from perusing solutions to the problems whether they have worked them out themselves or not. Students using this volume will still need to work out solutions of research questions using their own sources and adapting them to their own geographic locations, or using their own computational schemes, so this volume could well be useful for students in many course contexts. Enrichment material is included on the topics of some of the exercises. Advice for teachers who lack previous environmental experience, but who want to teach this material is also provided and makes it practical for such persons to offer a course based on these volumes.

Catalog Code: EVS/SA
150 pp., Paperbound, 1998, ISBN 0-88385-713-8
List: $18.95 MAA Member: $14.95

Phone in Your Order Now! ☎ *1-800-331-1622*

Monday – Friday 8:30 am – 5:00 pm FAX (301) 206-9789
or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

THE MATHEMATICAL ASSOCIATION OF AMERICA



A Course in Mathematical Modeling

Series: Classroom Resource Materials

Douglas Mooney and Randall Swift

Textbook

This book is intended as a text for a modeling course accessible to students who have mastered a one year course in calculus. It balances a variety of opposing modeling methodologies including theoretical models versus empirical models, analytical models versus simulation, deterministic models versus stochastic models, and discrete models versus continuous models. Most of the examples are drawn from real-world data or from models that have been used in various applied fields. The use of computers in both simulation and in mathematical analysis is an integral part of the presentation.

The authors emphasize the teaching of the modeling process as opposed to merely presenting models. They begin their book with the simple discrete exponential growth model, and successively refine it to include variable growth rates, multiple variables, growth rates fitted to data, and the effects of random factors. The last part of the book moves into continuous-time models. Issues of model validity and purpose are emphasized throughout.

Students taking a course based on this book should have some mathematical maturity, but will need little advanced knowledge. The book presents more advanced topics on an as-needed basis and serves to show how the different topics of undergraduate mathematics can be used together to solve problems. This perspective is valuable as either a road map for beginning students or as a capstone for more advanced students. The course presents elements of discrete dynamical systems, basic probability theory, differential equations, matrix algebra, stochastic processes, curve fitting, statistical testing, and regression analysis. Computer analysis is extensively used in conjunction with these topics.

You can also use this book if you are seeking applications to supplement a course in linear algebra, differential equations, difference equations, probability theory, or statistics.

Catalog Code: MML/JR
400 pp., Paperbound, 1999
ISBN 0-88385-712-X
List: $41.95   MAA Member: $32.95

Visit <www.wku.edu/~swiftrj/Modeling/ modeling.html> where you can visit the authors' website and download data sets, *Mathematica* files, and other modeling resources that execute the models described in the text.

Phone in Your Order Now! ☎ *1-800-331-1622*

Monday – Friday 8:30 am – 5:00 pm   FAX (301) 206-9789

or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

Shipping and Handling: Postage and handling are charged as follows: **USA orders (shipped via UPS):** $2.95 for the first book, and $1.00 for each additional book. **Canadian orders:** $4.50 for the first book and $1.50 for each additional book. Canadian orders will be shipped within 10 days of receipt of order via the fastest available route. We do not ship via UPS into Canada unless the customer specially requests this service. Canadian customers who request UPS shipment will be billed an additional 7% of their total order. **Overseas orders:** $3.50 per item ordered for books sent surface mail. Airmail service is available at a rate of $7.00 per book. Foreign orders must be paid in US dollars through a US bank or through a New York clearinghouse. Credit Card orders are accepted for all customers.

Name ______________________

Address ______________________

City ____________ State _____ Zip ________

Phone ______________________

QTY.	CATALOG CODE	PRICE	AMOUNT
______	MML/JR	______	______

All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

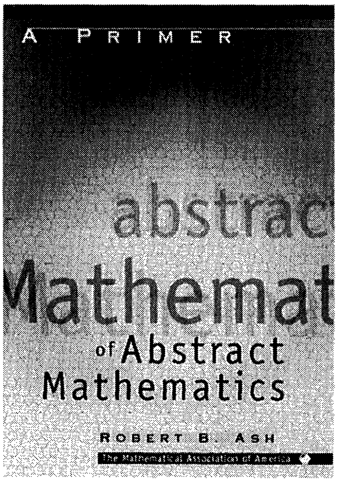
Shipping & handling ______

TOTAL ______

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ______________ Expires ___/___

Signature ______________________



THE MATHEMATICAL ASSOCIATION OF AMERICA

A Primer of Abstract Mathematics

Robert Ash

Series: Classroom Resource Materials

A Primer of Abstract Mathematics prepares the reader to cope with abstract mathematics, specifically abstract algebra. It can serve as a text for prospective mathematics majors, as well as for those students taking or preparing to take a first course in abstract algebra, or those in applied fields who need experience in dealing with abstract mathematical ideas.

Learning any area of abstract mathematics involves writing formal proofs, but it is equally important to think intuitively about the subject and to express ideas clearly and cogently. The author aids intuition by keeping proofs short and as informal as possible, using concrete examples which illustrate all the features of the general case, and by giving heuristic arguments when a formal development would take too long. The text can serve as a model on how to write mathematics for an audience with limited experience in formalism and abstraction.

Ash introduces several expository innovations in *A Primer of Abstract Mathematics.* He presents an entirely informal development of set theory that gives students the basic results that they will need in algebra. The chapter which presents the theory of linear operators introduces the Jordan Canonical Form right at the beginning, with a proof of existence at the end of the chapter.

Contents:

Logic and Foundations: Truth Tables, Quantifiers, Proofs, Sets, Functions, Relations.

Counting: Fundamentals, The Binomial and Multinomial Theorems, The Principle of Inclusion and Exclusion, Counting Infinite Sets.

Elementary Number Theory: The Euclidean Algorithm, Unique Factorization, Algebraic Structures, Further Properties of Congruence Modulo m, Linear Diophantine Equations and Simultaneous Congruences, Theorems of Euler and Fermat, The Möbius Inversion Formula.

Some Highly Informal Set Theory: Well-Orderings, Zorn's Lemma and the Axiom of Choice, Cardinal Numbers, Addition and Multiplication of Cardinals.

Linear Algebra: Matrices, Determinants and Inverses, The Vector Space F^n, Linear Independence and Bases, Subspaces, Linear Transformations, Inner Product Spaces, Eigenvalues and Eigenvectors.

Theory of Linear Operators: Jordan Canonical Form, The Minimal and Characteristic Polynomials, The Adjoint of a Linear Operator, Normal Operators, The Existence of the Jordan Canonical Form.

Appendix: An Application of Linear Algebra.

Catalog Code: BMA/JR
188 pp., Paperbound, 1998
ISBN 0-88385-709-X
List: $27.95 MAA Member: $21.95

Phone in Your Order Now! ☎ 1-800-331-1622

Monday – Friday 8:30 am – 5:00 pm FAX (301) 206-9789

or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

Shipping and Handling: Postage and handling are charged as follows: **USA orders (shipped via UPS):** $2.95 for the first book, and $1.00 for each additional book. **Canadian orders:** $4.50 for the first book and $1.50 for each additional book. Canadian orders will be shipped within 10 days of receipt of order via the fastest available route. We do not ship via UPS into Canada unless the customer specially requests this service. Canadian customers who request UPS shipment will be billed an additional 7% of their total order. **Overseas orders:** $3.50 per item ordered for books sent surface mail. Airmail service is available at a rate of $7.00 per book. Foreign orders must be paid in US dollars through a US bank or through a New York clearinghouse. Credit Card orders are accepted for all customers.

Name ______________________

Address ______________________

City ______________ State _____ Zip ________

Phone ______________________

QTY.	CATALOG CODE	PRICE	AMOUNT
_______	BMA/JR	_______	_______

All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

Shipping & handling _______

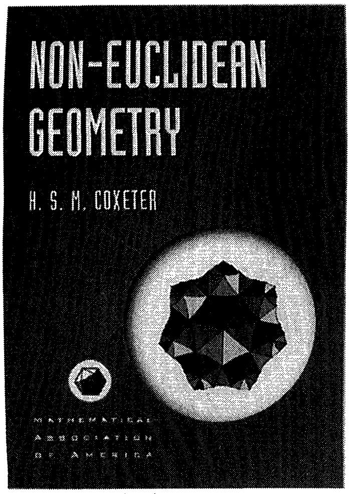
TOTAL _______

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ______________________ Expires ___/___

Signature ______________________

THE MATHEMATICAL ASSOCIATION OF AMERICA



Non-Euclidean Geometry

Sixth Edition

H. S. M. COXETER

Series: Spectrum

H. S. M. Coxeter's classic book on non-Euclidean geometry was first published in 1942, and enjoyed eight reprintings before it went out of print in 1968. The MAA is delighted to be the publisher of the sixth edition of this wonderful book, updated with a new section 15.9 on the author's useful concept of inversive distance.

Throughout most of this book, non-Euclidean geometries in spaces of two or three dimensions are treated as specializations of real projective geometry in terms of a simple set of axioms concerning points, lines, planes, incidence, order and continuity, with no mention of the measurement of distances or angles. This synthetic development is followed by the introduction of homogeneous coordinates, beginning with Von Staudt's idea of regarding points as entities that can be added or multiplied. Transformations that preserve incidence are called collineations. They lead in a natural way to elliptic isometries or "congruent transformations". Following a recommendation by Bertrand Russell, continuity is described in terms of order. Elliptic and hyperbolic geometries are derived from real projective geometry by specializing an elliptic or hyperbolic polarity which transforms points into lines (in two dimensions) or planes (in three dimensions) and vice versa.

An unusual feature of the book is its use of the general linear transformation of coordinates to derive the formulas of elliptic and hyperbolic trigonometry. The area of a triangle is related to the sum of its angles by means of an ingenious idea of Gauss. This treatment can be enjoyed by anyone who is familiar with algebra up to the elements of group theory.

Catalog Code: NEC/JR
320 pp., Paperbound, 1988
ISBN 0-88385- 522-4
List: $30.95 MAA Member: $24.50

Phone in Your Order Now! ☎ *1-800-331-1622*

Monday – Friday 8:30 am – 5:00 pm FAX (301) 206-9789

or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

Shipping and Handling: Postage and handling are charged as follows: **USA orders (shipped via UPS):** $2.95 for the first book, and $1.00 for each additional book. **Canadian orders:** $4.50 for the first book and $1.50 for each additional book. Canadian orders will be shipped within 10 days of receipt of order via the fastest available route. We do not ship via UPS into Canada unless the customer specially requests this service. Canadian customers who request UPS shipment will be billed an additional 7% of their total order. **Overseas orders:** $3.50 per item ordered for books sent surface mail. Airmail service is available at a rate of $7.00 per book. Foreign orders must be paid in US dollars through a US bank or through a New York clearinghouse. Credit Card orders are accepted for all customers.

QTY.	CATALOG CODE	PRICE	AMOUNT
____	NEC/JR	____	____

Name ____________

Address ____________

City ____________ State _____ Zip ______

Phone ____________

All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

Shipping & handling ______

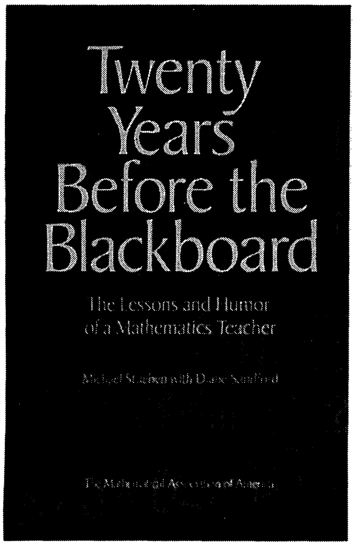
TOTAL ______

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ____________ Expires ___/___

Signature ____________

THE MATHEMATICAL ASSOCIATION OF AMERICA



Twenty Years Before the Blackboard

The Lessons and Humor of a Mathematics Teacher

Michael Stueben with Diane Sandford

Series: Spectrum

A perfect gift for the new teacher. . . or for anyone interested in the teaching of mathematics.

This book is the legacy of twenty years of mathematics teaching. During this time, the author searched for motivation techniques, mnemonics, insightful proofs, and serious applications of humor to aid his teaching. The result is this book: part philosophy, part humor, and part biography. Readers will be amused and enlightened on every page.

Mr. Stueben shows how he has used humor and wordplay to motivate his students. The book is filled with wonderful problems and proofs, as well as the author's insights about how to approach teaching problem solving to high school students. Sections of the book also treat the use of calculators and computers in the classroom. A section on mnemonics shows how teachers can use memory aids to help their students learn and retain material.

All in all, *Twenty Years Before the Blackboard* provides a goldmine of ideas for the classroom teacher. Although Mr. Stueben taught at the high school level, his book is an excellent "methods" book for mathematics teachers at all levels.

Read what Martin Gardner has to say about this fascinating book:

It's been decades since I read so entertaining a book about mathematics. The book is a treasure-trove of mathematical jokes, rhymes, anecdotes, word play, mnemonics, and beautiful proofs. For teachers there is an abundance of wise advice based on the author's twenty years in high school teaching. Mathematicians at all levels, from amateurs to college professors will not only chuckle over its gems, but learn much they did not know before. —Martin Gardner

Catalog Code: TYB/JR98
174 pp., Paperbound, 1998, ISBN 0-88385-525-9
List: $29.50 MAA Member: $23.50

Phone in Your Order Now! ☎ *1-800-331-1622*

Monday – Friday 8:30 am – 5:00 pm FAX (301) 206-9789

or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

Shipping and Handling: Postage and handling are charged as follows: **USA orders (shipped via UPS):** $2.95 for the first book, and $1.00 for each additional book. **Canadian orders:** $4.50 for the first book and $1.50 for each additional book. Canadian orders will be shipped within 10 days of receipt of order via the fastest available route. We do not ship via UPS into Canada unless the customer specially requests this service. Canadian customers who request UPS shipment will be billed an additional 7% of their total order. **Overseas orders:** $3.50 per item ordered for books sent surface mail. Airmail service is available at a rate of $7.00 per book. Foreign orders must be paid in US dollars through a US bank or through a New York clearinghouse. Credit Card orders are accepted for all customers.

Name ______________________

Address ______________________

City ____________ State _____ Zip ________

Phone ______________________

QTY.	CATALOG CODE	PRICE	AMOUNT
______	TYB/JR98	______	______

All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

Shipping & handling ______

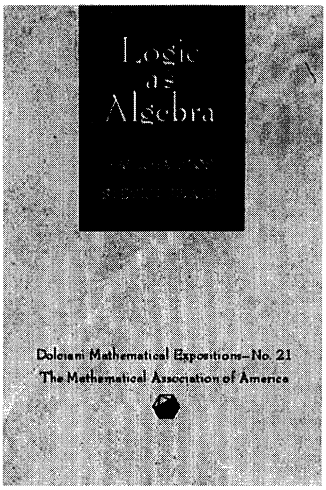
TOTAL ______

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ______________ Expires ___/___

Signature ______________________

THE MATHEMATICAL ASSOCIATION OF AMERICA



Logic as Algebra

Paul Halmos and Steven Givant

Series: Dolciani Mathematical Expositions

This book is based on the notes of a course in logic given by Paul Halmos. This book retains the spirit and purpose of those notes, which was to show that logic can (and perhaps should) be viewed from an algebraic perspective. When so viewed, many of its principal notions are seen to be old friends, familiar algebraic notions that were "disguised" in logical clothing. Moreover, the connection between the principal theorems of the subject and well-known theorems in algebra becomes clearer. Even the proofs often gain in simplicity.

Propositional logic and monadic predicate calculus—predicate logic with a single quantifier— are the principal topics treated. The connections between logic and algebra are carefully explained. The key notions and the fundamental theorems are elucidated from both a logical and algebraic perspective. The final section gives a unique and illuminating algebraic treatment of the theory of syllogisms—perhaps the oldest branch of logic, and a subject that is neglected in most modern logic texts.

The presentation is aimed at a broad audience—mathematics amateurs, students, teachers, philosophers, linguists, computer scientists, engineers, and professional mathematicians. Whether the reader's goal is a quick glimpse of modern logic or a more serious study of the subject, the book's fresh approach will bring novel and illuminating insights to beginners and professionals alike. All that is required of the reader is an acquaintance with some of the basic notions encountered in a first course in modern algebra. In particular, no prior knowledge of logic is assumed. The book could serve equally well as a fireside companion and as a course text.

Contents: **What is Logic?:** To count or to think; A small alphabet; A small grammar; A small logic; What is truth?; Motivation of the small language; All mathematics. **Propositional Calculus:** Propositional symbols; Propositional abbreviations; Polish notation; Language as an algebra; Concatenation; Theorem schemata; Formal proofs; Entailment; Logical equivalence; Conjunction; Algebraic identities. **Boolean Algebra:** Equivalence classes; Interpretations; Consistency and Boolean algebra; Duality and commutativity; Properties of Boolean algebras; Subtraction; Examples of Boolean algebras. **Boolean Universal Algebra:** Subalgebras; Homomorphisms; Examples of homomorphisms; Free algebras; Kernels and ideals; Maximal ideals; Homomorphism theorem; Consequences; The representation theorem. **Logic via Algebra:** Pre-Boolean algebras; Substitution rule; Boolean logics; Algebra of the propositional calculus; Algebra of proof and consequence. **Lattices and Infinite Operations:** Lattices; Non-distributive lattices; Infinite operations. **Monadic Predicate Calculus:** Propositional functions; Finite functions; Functional monadic algebras; Functional quantifiers; Properties of quantifiers; Monadic algebras; Free monadic algebras; Modal logics; Monadic logics; Syllogisms.

Catalog Code: DOL-21/JR98
152 pp., Paperbound, 1998, ISBN 0-88385-327-2
List: $27.00 MAA Member: $21.95

Phone in Your Order Now! ☎ 1-800-331-1622

Monday – Friday 8:30 am – 5:00 pm FAX (301) 206-9789

or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

Shipping and Handling: Postage and handling are charged as follows: **USA orders (shipped via UPS):** $2.95 for the first book, and $1.00 for each additional book. **Canadian orders:** $4.50 for the first book and $1.50 for each additional book. Canadian orders will be shipped within 10 days of receipt of order via the fastest available route. We do not ship via UPS into Canada unless the customer specially requests this service. Canadian customers who request UPS shipment will be billed an additional 7% of their total order. **Overseas orders:** $3.50 per item ordered for books sent surface mail. Airmail service is available at a rate of $7.00 per book. Foreign orders must be paid in US dollars through a US bank or through a New York clearinghouse. Credit Card orders are accepted for all customers.

Name ______________________

Address ______________________

City ______________ State _____ Zip ________

Phone ______________________

QTY.	CATALOG CODE	PRICE	AMOUNT
_______	DOL-21/JR98	_______	_______

All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

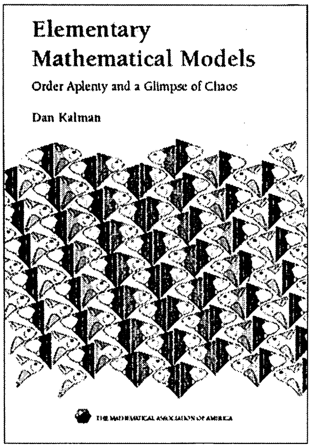
Shipping & handling ________

TOTAL ________

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ______________ Expires ___/___

Signature ______________________



THE MATHEMATICAL ASSOCIATION OF AMERICA

Elementary Mathematical Models

Order Aplenty and a Glimpse of Chaos

Series: Classroom Resource Materials

Dan Kalman

New in the Classroom Resource Materials Series...suitable as a text in college algebra, finite mathematics, precalculus, and liberal arts mathematics courses

Elementary Mathematical Models *(EMM)* claims a middle ground between college algebra and liberal arts mathematics. Like the college algebra course, *EMM* emphasizes the elementary functions of analysis: linear, quadratic, polynomial, and rational functions; square roots; exponentials and logarithms. These functions are the building blocks for the simple models that appear in first courses in the physical, life, and social sciences. And while *EMM* does not stress algebraic manipulation as an end in itself, it does recognize how important algebra is. Moreover, it provides students with the opportunity to see for themselves why algebra is needed, and what it contributes to formulating and analyzing models.

Like the liberal arts mathematics course, *EMM* makes a concerted effort to convey something of the scope, power, and fascination of mathematics, to students who may never study mathematics again. Each mathematical topic evolves naturally in formulating simple discrete models for inherently interesting contexts. For example, exponential functions emerge from the study of models exhibiting geometric growth—defined as growth that increases by equal proportions in equal periods of time. Throughout the course, a recurring theme is the evolution from simple recursive hypotheses (e.g., the population next year will be 10% greater than this year), to difference equations ($p_{n+1} = 1.1\,p_n$), to solutions ($p_n = p_0\,(1.1^n)$), to qualitative behavior of models. This theme appears repeatedly as the students encounter a series of increasingly sophisticated growth models, starting with arithmetic growth and ending with logistic growth. The course climaxes with an exploration of the chaotic behavior that can occur in logistic growth models.

The presentation is accessible even to students with a weak algebraic background. Throughout the book, numerical, graphical, and symbolic approaches are used systematically. There is a rich collection of examples and exercises. Reading comprehension exercises in each chapter provide a strong emphasis on reading and writing about mathematical concepts.

Contents: Overview; Sequences and Differences Equations; Arithmetic Growth; Linear Graphs, Functions and Equations; Quadratic Growth Models; Quadratic Graphs, Functions and Equations; Polynomials and Rational Functions; Fitting a Line to Data; Geometric Growth; Exponential Functions; More on Logarithms; Geometric Sums and Mixed Models: Logistic Growth; and Chaos in Logistic Models.

Catalog Code: EMM/JR
360 pp., 1997, Paperbound, ISBN 0-88385-707-3
List: $32.50 MAA Member: $25.95

Supplementary problem collection available with adoption orders. Call us toll free at 1-800-331-1622 for more details.

Phone in Your Order Now! ☎ 1-800-331-1622

Monday – Friday 8:30 am – 5:00 pm FAX (301) 206-9789

or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

Shipping and Handling: Postage and handling are charged as follows: **USA orders (shipped via UPS):** $2.95 for the first book, and $1.00 for each additional book. **Canadian orders:** $4.50 for the first book and $1.50 for each additional book. Canadian orders will be shipped within 10 days of receipt of order via the fastest available route. We do not ship via UPS into Canada unless the customer specially requests this service. Canadian customers who request UPS shipment will be billed an additional 7% of their total order. **Overseas orders**: $3.50 per item ordered for books sent surface mail. Airmail service is available at a rate of $7.00 per book. Foreign orders must be paid in US dollars through a US bank or through a New York clearinghouse. Credit Card orders are accepted for all customers.

Name ______________________

Address ______________________

City ______________ State _____ Zip _______

Phone ______________________

QTY.	CATALOG CODE	PRICE	AMOUNT
______	EMM/JR	______	______

Shipping & handling ______

All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

TOTAL ______

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ______________ Expires ___/___

Signature ______________________

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036